

Tesis Doctoral

Estudio de los determinantes moleculares que regulan la eficiencia de N-glicosilación de proteínas

Couto, Paula Monserrat

2018-03-26

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Couto, Paula Monserrat. (2018-03-26). Estudio de los determinantes moleculares que regulan la eficiencia de N-glicosilación de proteínas. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Couto, Paula Monserrat. "Estudio de los determinantes moleculares que regulan la eficiencia de N-glicosilación de proteínas". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2018-03-26.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Química Biológica

Estudio de los determinantes moleculares que regulan la eficiencia de *N*-glicosilación de proteínas.

Tesis presentada para optar por el título de Doctor de la Universidad de Buenos Aires en el área de Química Biológica.

Lic. Paula Monserrat Couto

Director de tesis: Dr. Julio Javier Caramelo

Consejero de estudios: Dr. Juan Carlos Calvo

Lugar de trabajo: Laboratorio de Biología Estructural y Celular. Fundación Instituto Leloir e IIBBA-CONICET, Ciudad Autónoma de Buenos Aires, Argentina.

Ciudad Autónoma de Buenos Aires, año 2018

ÍNDICE

RESUMEN	4
ABSTRACT.....	5
AGRADECIMIENTOS	6
INTRODUCCIÓN.....	9
I. Flujo de la información genética.....	9
II. Uso de codones.....	9
III. Velocidad de traducción.....	11
IV. Codones raros.....	14
V. Estructura del ARNm	16
VI. <i>N</i> -glicosilación.	16
HIPÓTESIS Y OBJETIVOS.....	23
VII. Hipótesis de trabajo.	23
VIII. Objetivos de trabajo.	24
MATERIALES Y MÉTODOS.....	26
IX. ADN, clonado y mutaciones.	26
X. Cepa bacteriana y medio de cultivo empleado.	27
XI. Transformación bacteriana	27
XII. Extracción plasmídica	27
XIII. Generación de secuones	28
XV. Medios de cultivo.....	31
XVI. Revitalización de líneas celulares.	31
XVII. Mantenimiento de líneas celulares.	31
XVIII. Criopreservación de líneas celulares.	31
XIX. Transfección.	31
XX. Lisis celular.....	32
XXI. Marcación isotópica celular.	32

XXII. Inmunoprecipitación.....	33
XXIII. Microscopía.....	33
XXIV. Geles.....	33
XXV. Revelado.....	34
XXVI. Western Blot.....	34
XXVII. Clonado y expresión de proteínas en <i>E. coli</i>	35
XXVIII. Análisis espectroscópicos.....	36
XXIX. Simulaciones de dinámica molecular.....	37
RESULTADOS.....	38
Diseño de la proteína de trabajo.....	38
Clonado del ADN en vectores de expresión en células de mamífero.....	40
Mutantes de <i>N</i> -glicosilación.....	43
HEK293 KNOCKOUT.....	54
ENSAYOS DE VELOCIDAD DE SÍNTESIS PROTEICA.....	61
EL RETORNO DE LA C70.....	64
EXPRESIÓN Y PURIFICACIÓN DE LAS <i>sfGFPs</i>	68
SIMULACIONES DE DINÁMICA MOLECULAR.....	82
DISCUSIÓN.....	86
USO DE CODONES.....	86
VELOCIDAD DE SÍNTESIS.....	91
HEK293 KNOCKOUT.....	92
INFLUENCIA DE LA IDENTIDAD DEL RESIDUO 70.....	97
ANEXO.....	103
TABLA DE USO DE CODONES EN HUMANOS.....	103
Secuencias nucleotídicas finales de las <i>cf-sfGFP</i> no glicosiladas.....	106
SECUENCIAS NUCLEOTÍDICAS DE LAS MUTANTES DE <i>N</i> -GLICOSILACIÓN.....	108
BIBLIOGRAFÍA.....	115

Estudio de los determinantes moleculares que regulan la eficiencia de *N*-glicosilación de proteínas

RESUMEN

La *N*-glicosilación es una de las modificaciones post-traduccionales más frecuentes en proteínas de eucariotas. Cerca de un cuarto de las mismas son *N*-glicosiladas al ingresar al lumen del retículo endoplásmico (RE) a través del translocón SEC61. Dicha modificación es catalizada por la enzima oligosacariltransferasa (OST), un complejo oligomérico que se asocia al SEC61. La OST transfiere un oligosacárido de alta manosa a la cadena lateral de asparaginas que se encuentran en el contexto de ASN-X-SER/THR (donde X no puede ser prolina). A esta secuencia consenso se le conoce como secuón de *N*-glicosilación.

Los *N*-glicanos cumplen diversos roles biológicos. Por ejemplo, asisten al plegado proteico en el RE y pueden participar en numerosos eventos de reconocimiento molecular. La presencia de un secuón de *N*-glicosilación no garantiza su ocupación por la OST. Por esta razón, una proteína puede presentarse como una mezcla de secuones parcialmente ocupados, un fenómeno conocido como macroheterogeneidad de la *N*-glicosilación. Se sabe que algunos secuones no son un buen sustrato para la OST, por ejemplo, aquellos que presentan un residuo ácido en el medio. Sin embargo, los determinantes moleculares para tal comportamiento son desconocidos en gran medida. Aquí se estudia cómo la dinámica de biosíntesis y plegado modulan este proceso complejo. Por una parte, se encontró que los codones sinónimos en los genes que codifican a las glicoproteínas no afectarían apreciablemente a la eficiencia de *N*-glicosilación. Por otra parte, se pudo determinar el efecto de la estabilidad termodinámica de las proteínas sobre la eficiencia de la *N*-glicosilación post-traducciona.

PALABRAS CLAVE: *N*-glicosilación, macroheterogeneidad, oligosacariltransferasa, secuón, plegado

Study of the molecular determinants that regulate the efficiency of protein *N*-glycosylation

ABSTRACT

N-glycosylation is one of the most frequent protein modifications in eukaryotes. Nearly one quarter of eukaryotic proteins are *N*-glycosylated as they enter into the lumen of the endoplasmic reticulum (ER) through the SEC61 translocon. This modification is carried out by the oligosaccharyltransferase enzyme (OST), an oligomeric complex associated with the SEC61. OST transfer a high mannose glycan to the lateral chain of ASN residues within the context ASN-X-SER/THR (where X cannot be PRO). This consensus sequence is known as *N*-glycosylation sequon.

N-glycans fulfill several biological roles. For example, they assist protein folding in the ER and they can participate in several molecular recognition events. The presence of an *N*-glycosylation sequon does not guarantee its full occupation by the OST. For this reason, a protein may present a mixture of partially occupied sequons, a phenomenon known as *N*-glycosylation macroheterogeneity. It is known that some sequons are poorly recognized by the OST, for instance those displaying acidic residues in the middle. Nevertheless, the molecular determinants for this behavior are largely unknown. Here we study how the protein biosynthesis dynamics and folding modulate this complex process. We showed that codon usage does not affect appreciably the efficiency of *N*-glycosylation. In addition, we could determine the effect of protein's thermodynamic stability on the efficiency of post-translational *N*-glycosylation.

KEY WORDS: *N*-glycosylation, macroheterogeneity, oligosaccharyltransferase, sequon, folding

AGRADECIMIENTOS

Llegar a esta instancia no se logra con tan solo cinco años. Este proceso comienza desde mucho antes y trataré de no olvidarme de mencionar a nadie. Comenzaré por agradecerle a mi querida escuela, el ENSAM de Banfield, por brindarme una educación pública de excelencia. Allí me encontré con profesores que nos motivaban a estudiar más allá de la currícula, incentivándonos a participar en certámenes como las olimpiadas de química, matemática y biología. Debo destacar los nombres de Mirtha López y Adriana Altieri quienes me hicieron amar la biología y me ayudaron a definir que eso era lo que quería estudiar.

Ya en la Facultad de Ciencias Exactas y Naturales de la UBA, tuve el enorme placer de cruzarme con excelentes docentes, no sólo por sus enseñanzas académicas sino por ser además, excelentes personas. Las peores materias podían sobrellevarse con su apoyo y las sonrisas; sobre todo en los laboratorios y clases de problemas. Mil gracias a Omar Coso, Luciana Balboa, Analía Silva, Matías Jobbagy, Pedro Aramendía, Virginia Diz, Pablo Hoijemberg, Ignacio Ruiz, Mati Galetti, Mariano Galvagno, Claudio Chillote, Leo Vanni, Nino Cafure, Daniel Cataldo, David Bilenca, Luis Diorio, Gabriel Manrique, Martín Berón de Astrada, Sol Fustiñana, Fernando Alonso, Adriana Cirigliano, Guillermo Noriega, Charlie Guardia, Julio Caramelo, Juan Carlos Calvo, Leandro Miquet, Fernando Aprile, Alf Soler Bistué, Norberto Iusem, quienes fueran tan especiales en mi paso por la FCEyN.

A mis compañeros de la facu con quienes compartimos tantas cursadas, tantas fiestas y, más importante aún, tantos partidos: a “Los Pinos”, a “Los Cueva”, a “Más que una cara bonita”, a “Choripete”, a Yami Juárez, Yami Agro, Caro Flumián, Anita Favier, Anto Merlotti, Marce Villarreal, Darío Puebla, Diego Rodríguez, Guille Lobo, Luli Rabitti, Noe Fiumara, Romi Maltaneri, Nanu, Nacho Aldaya, Adrián Salatino, Caro Oliveto, Luis De Haro, Days Fornés, Nico Bujak, Benja Basanta, Pato Pavía, Pepe Ares, Willy Turazzini, Damián Pérez, Flor Voda y Nil Dombrovsky.

Quienes me conocen saben lo necesario que es el fútbol en mi vida. Por eso quiero agradecerle a Diego Bruscagin por permitirme jugar en el torneo interno masculino en el Pommies y por haberme dado un lugar en el equipo de la UBA, donde pude descargar las tensiones, divertirme y aprender junto a buenas compañeras.

Al club Boca Jrs, que me abrió las puertas para conocer el futsal y al club Pinocho que junto con sus entrenadores y pinochas me alegran día a día.

A los Juegos Deportivos del CONICET, donde conocí cientos de amigos: especialmente a mis Lobas, Changas, a Fran y Chelito, a Nico Lavagnino, a Nico Moreyra y a Tito Costa. A los choris por hacer de los mixtos una fiesta. A Alfio Ruiz, el mejor entrenador de boxeo.

Al Leloir, un instituto repleto de gente hermosa. Diego Matías Álvarez, Changui Iezzi, Colo Bichara, Lu Gutiérrez, Dora, Lean Güttlein, Dani Chirico, Caffé, Santi W, Elvia Rivas, Sabri Mansilla, Iván Stigliano, Nati Contreras, Marce Pérez-Pepe, a nuestro querido vecino Pablo Cerdán, a Ángeles Zorreguieta, a todo el lab 209, a Maxi Sánchez, a las chicas de fútbol, a toda la gente de mantenimiento, biblioteca, soporte informático, limpieza y administración.

A Marta Bravo y a Susana Raffo por estar siempre a disposición.

Al lab 302 por compartir reactivos, buenos momentos y, sobre todo, ¡por las excelentes lentejeadas del Dr Parodi!

A Eugenia Clerico por ser nuestra dealer.

A mi pana, Jorge Yanez, por ser un hermano de oro y darme un sobrino tan bello.

A Facundo Cledón por estar a mi lado desde primer grado (aunque haya un río de por medio).

A Gaby Gómez, Lucas Landolfo y Máx López Medus por enseñarme a trabajar en el labo.

Al polaco Goyeneche.

Al Dr. Reid Gilmore por enviarnos las líneas celulares que le darían vida a esta tesis.

A SciHub y a los miembros de BPPF por cada paper que necesité.

A mi ídolo de la FCEyN, con quien soñaba trabajar y que me abrió los brazos cuando había perdido las esperanzas. Gracias Julio por ser un ejemplo a seguir, por ser como un padre, por aguantar mi mala onda y pocas palabras, por estar cuando te necesité, por hacer tan divertido ir al laboratorio, por pipetear por mí y por dejarme seguir siendo parte del lab como posdoc.

Al grupo de liberen el c....que siempre estuvieron para levantarme cuando caí, que me hicieron llorar de la risa en innumerables ocasiones. Gracias Vane Mattera, Gio Gallo, Chris Lorenzo y Anita Diez por cada almuerzo, por cada chat.

Gracias Sole por enseñarme tu léxico tucumano, por abrirme las puertas de tu casa, por ser tan buena persona.

Gracias a mis inmejorables compañeros de viaje: a mi "Sr" Rami Orsi y a mi falso padre, Carlitos Labriola.

A Charlie Guardia por estar tan presente, por ayudarme a la distancia, por tanto cariño.

A Lean Vargas por quererme tanto, ayudarme y ser de oro.

Al CONICET por haberme dado esta oportunidad y seguir apostando en este viaje.

A mis doctores, Charlie Lemessof y Norberto Traverso por cuidarme y ayudarme a recuperarme pronto. A Ale Fernández Novoa, que siempre me salva cuando tengo una urgencia.

A Alicia que siempre me impulsaba a entrar en un laboratorio y a Pau Arroyo que contribuyó en lo mismo.

A mi familia lomense y marplatense. A la tía Gabriela que siempre me prestaba la impresora para los resúmenes, a Cristi que me prestaba su casa para dormir, a Gonzalo que siempre está para ayudar, a Facu (que cuando no molesta y se cansa de que le insista) me hace favores, a mis viejos que me bancaron toda la carrera para que no tuviera que trabajar. Al pela, a los locos de mis sobris y cuñad@s.

A todos los que me ayudaron y apoyaron emocionalmente cuando durante la finalización de este escrito, muy inoportunamente, me rompí la rodilla y el corazón en mil pedazos.

A los miembros de mi comité de seguimiento de tesis: Fernanda Ceriani, Vanesa Gottifredi, Ricardo Wolosiuk.

A los jurados, por aceptar ser parte de esto.

INTRODUCCIÓN

I. Flujo de la información genética.

En 1958, Francis Crick enunció el dogma central de la biología molecular, donde se establece el flujo de la información genética (**figura 1**). Las células, mediante el proceso de transcripción y traducción, producen una proteína compuesta por una cadena de aminoácidos, estando establecidas en el código genético las reglas básicas para la conversión de un gen en proteína.

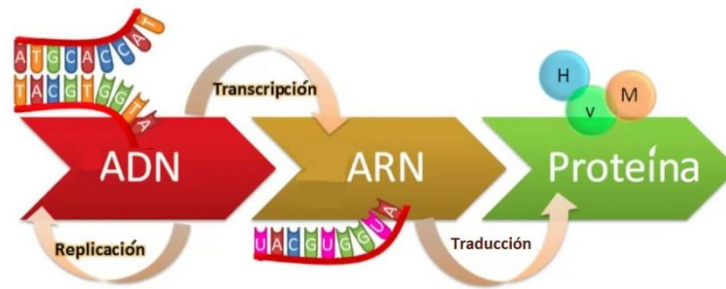


Figura 1. Dogma central de la biología molecular: el ADN es transcrito a una molécula de ARN mensajero, la cual será traducida para generar una proteína.

Las proteínas constituyen más de la mitad del peso seco de una célula, siendo su síntesis primordial para el mantenimiento, crecimiento y desarrollo celular. El proceso, conocido como traducción, es mediado por ribosomas en tres etapas. Durante la iniciación, el ribosoma se posiciona sobre el codón de inicio (AUG) junto con el ARN de transferencia (ARNt) correspondiente. En la elongación, se va construyendo la cadena polipeptídica siguiendo el marco de lectura del ARN mensajero (ARNm). Los aminoácidos son entregados al ribosoma mediante ARNt que componen complejos ternarios y que sirven de moléculas adaptadoras entre el aminoácido y el codón ubicado en el sitio A del ribosoma. La traducción finaliza cuando el ribosoma alcanza un codón de terminación (“stop”), liberando al polipéptido y desensamblándose las subunidades ribosomales¹. Múltiples proteínas pueden ser sintetizadas en simultáneo sobre una molécula de ARNm, conformando una estructura conocida como polisoma, que consiste en varios ribosomas traduciendo al mismo mensajero.

II. Uso de codones.

Los codones son triplete de nucleótidos en el ARNm compuestos por 4 bases: adenina (A), citosina (C), guanina (G) y uracilo (U). Ello permite la generación de 64 posibles codones, tres de los cuales representan señales de terminación de la traducción. Los restantes 61 codifican a los 20 posibles aminoácidos, resultando en una redundancia donde múltiples

codones pueden codificar al mismo aminoácido (**figura 2**). Los únicos residuos que no presentan redundancia son metionina y triptófano, los cuales son codificados respectivamente por los codones AUG y UGG. La redundancia del código genético ofrece una oportunidad a la evolución para sintonizar la eficiencia y precisión de la producción proteica a varios niveles, manteniendo la misma secuencia aminoacídica².

En cada organismo no hay una distribución al azar sino que existe una preferencia por ciertos codones por sobre otros, fenómeno conocido como sesgo en el uso de codones³. Así, los codones sinónimos aparecen con diferentes frecuencias y su ocurrencia no sólo varía entre especies, sino también dentro de genes del mismo genoma. Las preferencias son más fuertes en genes de alta expresión a través de organismos diversos, sugiriendo la existencia de una presión de selección que permitiría el uso eficiente del aparato traduccional durante la síntesis de proteínas abundantes⁴.

	U		C		A		G		
U	UUU	Fenilalanina	UCU	Serina	UAU	Tirosina	UGU	Cisteína	U C A G
	UUC		UCC		UAC		UGC		
	UUA	Leucina	UCA		UAA	Stop	UGA	Stop	
	UUG		UCG		UAG		UGG	Triptófano	
C	CUU	Leucina	CCU	Prolina	CAU	Histidina	CGU	Arginina	U C A G
	CUC		CCC		CAC		CGC		
	CUA		CCA		CAA	Glutamina	CGA		
	CUG		CCG		CAG		CGG		
A	AUU	Isoleucina	ACU	Treonina	AAU	Asparagina	AGU	Serina	U C A G
	AUC		ACC		AAC		AGC		
	AUA		ACA		AAA	Lisina	AGA	Arginina	
	AUG	Metionina	ACG		AAG		AGG		
G	GUU	Valina	GCU	Alanina	GAU	Ácido Aspartico	GGU	Glicina	U C A G
	GUC		GCC		GAC		GGC		
	GUA		GCA		GAA	Ácido Glutámico	GGA		
	GUG		GCG		GAG		GGG		

Figura 2. Codones del ARNm y su correspondencia aminoacídica.

Muchas células contienen menos de 61 tipos de ARNt, lo cual se explica por la capacidad de algunos ARNt de decodificar a más de un codón. Esto es posible debido a que existen dos posibles maneras de decodificación: **(1)** vía estricto apareamiento de bases por Watson-Crick (WC) en las tres posiciones de la interacción codón:anticodón, ó **(2)** a través de apareamiento

del tipo no-Watson-Crick en la tercera posición del codón (conocida como interacción de tipo “wobble”)⁵. Para el “wobbling” existen reglas que se muestran en la **tabla 1**.

base 3' del codón	base 5' del anticodón
G	C
C, U	G
A	U
A, C, U	I

Tabla 1. Reglas de apareamiento por wobbling en eucariotas. La inosina (I) es una modificación frecuente de la adenosina que ocurre en la posición wobble de muchas especies de ARNt.

III. Velocidad de traducción.

El sesgo en el uso de codones es una característica universal de los genomas procariotas y eucariotas y, se ha propuesto que regula la eficiencia de traducción, la precisión y el plegado proteico, basándose en la asunción de que el uso de codones afecta la dinámica de traducción⁶. Que haya selección en el uso de codones podría deberse a la necesidad de maximizar la velocidad de elongación y aumentar la concentración de ribosomas libres, y/o a minimizar la incorporación de aminoácidos incorrectos en la cadena naciente⁷. Una teoría aceptada que explica la significancia biológica de la no uniformidad en el uso de codones es que los mismos son traducidos a diferentes velocidades^{8,9}. Así, es posible cambiar la tasa de emergencia del polipéptido desde el ribosoma mediante la elección de los codones.

La síntesis proteica es un proceso costoso en cuanto a energía y recursos celulares¹⁰. El proceso de selección natural sobre el sistema tiende a favorecer una máxima eficiencia y fidelidad, minimizando en paralelo el gasto energético, siendo un ejemplo de esto el uso de codones. Dicho sesgo se cree que reduce aproximadamente un 5% el requerimiento de los ribosomas que si se usaran todos los codones por igual¹¹.

Tradicionalmente se han usado diferentes criterios para predecir la velocidad de traducción de los codones, incluyendo su frecuencia de uso y la concentración de especies de ARNt capaces de decodificarlos, lo cual no siempre se correlaciona. Los ARNt correspondientes a codones sinónimos pueden variar en cantidad dentro de las células¹².

Si se considera a los ARNt como la oferta durante la traducción, luego, el uso de codones en el transcriptoma sería la demanda¹³. El balance entre la oferta y demanda (el grado de

adaptación del pool de ARNt al uso de codones en el transcriptoma) puede afectar los niveles de producción de proteínas. ¿Por qué las células habrían de coordinar la oferta y demanda en la traducción? La coordinación aseguraría la alta expresión de ciertos genes; así, genes que debieran ser expresados en altos niveles en un dado estado celular, serían traducidos más eficientemente debido a una buena adaptación al pool de ARNt en ese estado, siendo un ejemplo de ello el caso de *Drosophila*, donde los genes de alta expresión son codificados principalmente por codones que se corresponden con ARNt abundantes, permitiendo una traducción rápida y de alta fidelidad¹⁴. Otra visión implica que la coordinación podría permitir la optimización de la asignación de recursos de la traducción en la célula. De acuerdo a esto último, cuando la oferta y la demanda están balanceadas, se espera que los ribosomas fluyan sin congestión, minimizando los costos¹³. Mediante la expresión de ARNt en proporción al sesgo en sus codones se puede obtener una alta tasa de traducción de los mensajeros, con una inversión relativamente baja del aparato traduccional por masa

La existencia de una relación entre el uso de codones y la abundancia del anticodón, donde el uso de codones está adaptado para coincidir con el anticodón de los ARNts, fue demostrada por vez primera en los comienzos de la década del 1980 en *E. coli*^{15,16}. El sesgo en el uso de codones correlaciona con los niveles de ARNt en *E. coli*¹⁵, *Saccharomyces cerevisiae*^{17,18}, *Saccharomyces pombe*¹⁹, *Neurospora crassa*²⁰, *Caenorhabditis elegans*²¹, *Drosophila spp*²²., *Mycoplasma capricolum*²³ y *Arabidopsis thaliana*²⁴. Con la terminación de la secuenciación del genoma de *C. elegans*, se reportó que el número de genes para ARNt y su uso de codones en el genoma están co-adaptados para una traducción óptima en genes de alta expresión²¹. Sin embargo, existen excepciones, como el caso de algunos aminoácidos en cloroplastos de tabaco, donde se ha determinado que no hay correlación entre la abundancia del codón y la de su ARNt correspondiente²⁵. En el caso de los humanos, algunos estudios no han encontrado evidencia de selección traduccional en humanos (adaptación del uso de codones al pool de ARNt), sugiriendo que los codones sinónimos en esa especie no serían seleccionados para maximizar la eficiencia traduccional²⁶. Por otra parte, otros estudios han señalado una selección traduccional débil pero significativa en humanos²⁷.

La tasa en la que un ribosoma agrega un aminoácido codificado por un codón en particular a la cadena naciente puede variar significativamente y se le conoce como tasa de traducción del codón²⁸. Los codones frecuentes han sido tradicionalmente considerados rápidos, mientras que los raros han sido predichos como lentos. Así, en la mayoría de los casos el término codón raro significará al mismo tiempo que se traduce lentamente²⁹.

Se ha propuesto que la tasa de elongación por codón depende del pool de ARNt y que el ribosoma debe esperar más tiempo para que un ARNt poco frecuente ingrese al sitio A³⁰. Un ARNt *near-cognate* (NC) se define como aquel que puede establecer un apareamiento de WC con los dos primeros nucleótidos del codón pero que tiene cargado un aminoácido distinto al codificado por el codón. El proceso de muestreo y expulsión de un ARNt NC impone un costo cinético sustancial en los casos donde los ARNt *cognate* (aquellos que tienen cargado el aminoácido correcto) son poco abundantes en comparación con los NC. Generalmente se asume que un aumento de la relación NC sobre cognate disminuye la velocidad de traducción como resultado de una competencia en el ribosoma³¹. Existe evidencia en bacteria de que un exceso de ARNt *cognate* sin aminoácido cargado puede competir por el ingreso al sitio A del ribosoma con el complejo ternario, disminuyendo así la tasa de elongación durante la síntesis proteica³².

Los codones raros tienen el potencial de bajar la tasa de elongación traduccional debido a tiempos de permanencia relativamente altos del ribosoma en su búsqueda por ARNt raros³³. En su trabajo Spencer *et al.*³⁴ mostraron que cuando una proteína es codificada por codones poco abundantes, la misma muestra menores niveles de expresión debido a una menor velocidad de traducción respecto de la secuencia silvestre, y vieron el fenómeno opuesto en el caso de la secuencia codificada por los codones más abundantes (que correlacionan con los ARNt más abundantes).

En su estudio, Sørensen *et al.*³⁵, realizaron mediciones *in vivo* y mostraron que la inserción de codones raros o frecuentes afecta a la tasa de elongación, siendo la tasa de incorporación de aminoácidos en los codones frecuentes seis veces más rápida que en los raros. Sin embargo, también observaron que dos codones sinónimos leídos por el mismo ARNt fueron traducidos con una tasa que difiere tres veces, implicando que la traducción de esos codones no se debe únicamente a la abundancia de ARNt.

Los perfiles de ribosomas (“ribosome profiling”) generan un set de fragmentos de ARNm protegidos por ribosomas (huellas, “footprints”), que muestran la densidad ribosomal a lo largo de los ARNm con una resolución a nivel de codón individual. El estancamiento de ribosomas debido a una traducción lenta genera una acumulación de huellas ribosomales en posiciones específicas dentro de un gen y, en consecuencia, aparece un pico de densidad ribosomal. Analizando perfiles ribosomales de *E. coli*, *B. subtilis*, *C. elegans* y *S. cerevisiae*³⁶ se vio por vez primera que existe una correlación significativa entre las concentraciones de ARNt y el tiempo de decodificación de un codón. Mientras que en otros estudios no se halló correlación entre una

alta tasa de elongación y el uso de codones frecuentes³⁷. Tampoco se ha encontrado una correlación entre el tiempo de decodificación del codón y el número de copias génicas de ARNt (tAI) en perfiles ribosomales de ratón³⁶.

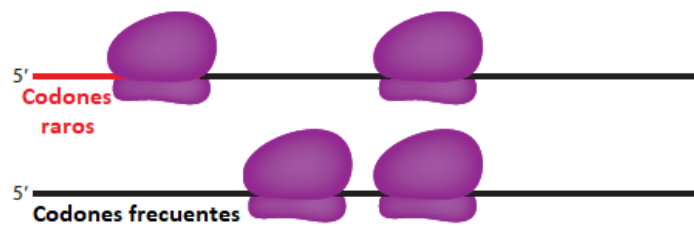
Ciertos estudios han sugerido que la velocidad de decodificación puede variar según sea por apareamiento de WC ó vía wobbling, resultando en tasas más lentas con el decodificado de tipo wobble^{22, 34, 38}. Usando perfiles ribosomales, en un trabajo se obtuvo la ocupación ribosomal *in vivo* como una función de la identidad del codón en *C. elegans* y en células HeLa, donde se vió una gran ocupación en codones que son traducidos vía wobbling G:U en comparación con sus codones sinónimos traducidos por interacciones G:C. Estos resultados apoyan a un modelo en donde la translocación ribosomal resulta enlentecida en posiciones con wobbling³⁸.

IV. Codones raros

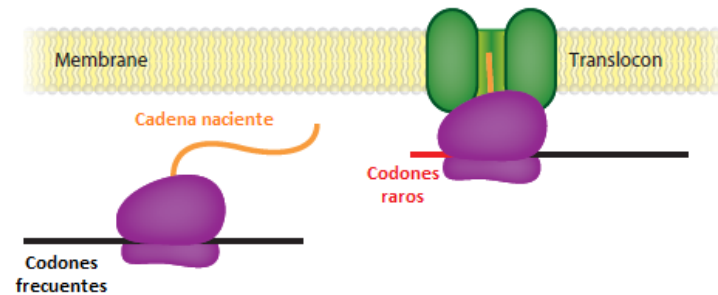
Se ha visto que los codones poco frecuentes o raros no están distribuidos al azar en las secuencias codificantes, sino que tienden a aparecer en *clusters* en la mayoría de las especies. Esto llevó a la hipótesis de que se encuentran bajo una presión de selección positiva en ciertas locaciones, debido a algún rol funcional³⁹.

Existe un patrón común que predomina en genes de alta expresión en eucariotas y procariotas: una extensión de 30-50 codones relativamente raros luego del codón de inicio. Se observó que el extremo 5' de las secuencias codificantes de proteínas de secreción está más enriquecido en codones raros que el de las proteínas citosólicas y, en consecuencia, se han planteado tres posibles explicaciones⁴⁰ (**figura 3**). Una posibilidad es que ello contribuya a aumentar el direccionamiento a la membrana del RE y la eficiencia de secreción. Otra, es que debido a que los codones raros reducen la tasa de traducción, crearían una rampa que optimizaría el espaciamiento entre ribosomas sobre el ARNm, previniendo así los embotellamientos de ribosomas. Por último, se ha propuesto que podría ser un subproducto de la selección actuando sobre el contenido GC, ya que un alto porcentaje GC suele generar estructuras de ARNm más estables y ello podría interferir con el inicio de la traducción⁴¹. Por otra parte, dado que el ribosoma procariota tiene actividad helicasa, se espera que sea menos susceptible al atascamiento inducido por estructuras en el ARNm de lo que son los eucariotas.

a Hipótesis de la rampa: espaciamiento ribosomal



b Hipótesis de secreción: pausa traduccional para direccionamiento



c Hipótesis de la estructura del ARNm: inicio de traducción

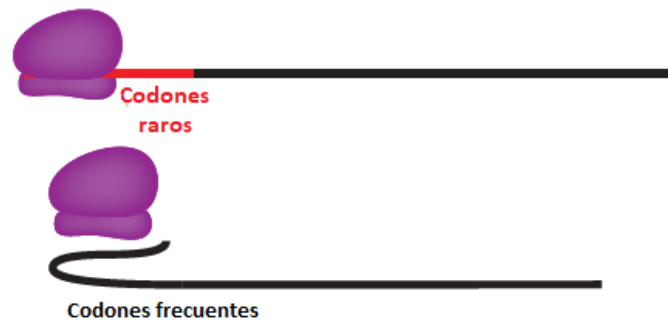


Figura 3. Diferentes hipótesis acerca de la ocurrencia de codones poco frecuentes en el extremo 5' del ARNm. Modificada de "Roles for Synonymous Codon Usage in Protein Biogenesis".

Dicha rampa de codones raros quizá explicaría por qué la expresión de genes foráneos a veces falla, ya que, debido a cambios en la disponibilidad de los ARNt, la rampa desaparecería en un sistema heterólogo. También ayudaría a explicar por qué la sobreexpresión de ARNt poco frecuentes (que mejoraría la producción proteica) no siempre funciona y que incluso puede ser contraproducente: la sobreexpresión excesiva de un ARNt poco abundante podría significar cambios en la eficiencia de traducción de ciertos codones, quebrantando así la rampa en detrimento de la síntesis proteica⁴².

Cannarozzi⁴³ y colegas observaron que codones idénticos tienen una fuerte tendencia a ser utilizados nuevamente cuando un aminoácido se repite, y que, si el mismo codón no es reusado, existe una tendencia a utilizar el codón sinónimo más cercano del wobbling. A tal ocurrencia se le llamó autocorrelación. Existe una menor densidad ribosomal en ARNm reporteros con codones autocorrelacionados que en sus contrapartes con codones

anticorrelacionados (la densidad ribosomal es inversamente proporcional a la tasa de elongación). Se concluye que la ventaja de reusar un codón proviene de la posibilidad de reutilizar el ARNt correspondiente, lo cual implica que los ARNt que salen del ribosoma permanecen asociados con la maquinaria traduccional, estando disponibles cuando el próximo codón idéntico aparece. Este modelo requiere que las enzimas responsables de unir el aminoácido a los ARNt (aminoacil-ARNt sintetetas) estén también asociadas a los ribosomas, de lo cual hay evidencias⁴⁴.

V. Estructura del ARNm

La secuencia nucleotídica primaria determina la estructura secundaria de los ARNm. Considerando que la formación de estructura secundaria en los mensajeros se basa en apareamientos nucleotídicos, un cambio en la secuencia puede alterarla considerablemente.

Estructuras estables en el ARNm pueden generar pausas traduccionales, pudiendo proveer de un mecanismo adicional para regular la velocidad de traducción. Estructuras plegadas en la región codificante representan barreras cinéticas que disminuyen la tasa de elongación peptídica, debido a que el ribosoma debe romper las estructuras que encuentra en su camino para translocar al siguiente codón⁴⁵. Se ha hipotetizado que genes que no se expresan mucho favorecen la formación de estructuras en el ARNm y que, a la inversa, una alta expresión proteica es facilitada por una organización estructural débil alrededor del 5' del ARNm⁴⁶.

La eficiencia traduccional estaría entonces determinada por el bias en el uso de los codones, el contenido GC y la estructura secundaria del ARNm.

VI. N-glicosilación.

El retículo endoplasmático (RE) es el compartimiento subcelular involucrado en la síntesis, modificación y plegado de proteínas que serán destinadas a secreción, membrana plasmática o distintas organelas de los sistemas endocítico y exocítico. Las proteínas destinadas a entrar en la vía secretora usualmente poseen una extensión en el N-terminal llamada péptido señal. Dicha secuencia señal permite el transporte de la proteína nascente hacia sitios de translocación sobre la membrana del RE, donde luego será removida proteolíticamente⁴⁷. A medida que son sintetizadas, estas proteínas ingresan al lumen del RE mediante un complejo de translocación denominado Sec61 (**figura 4**).

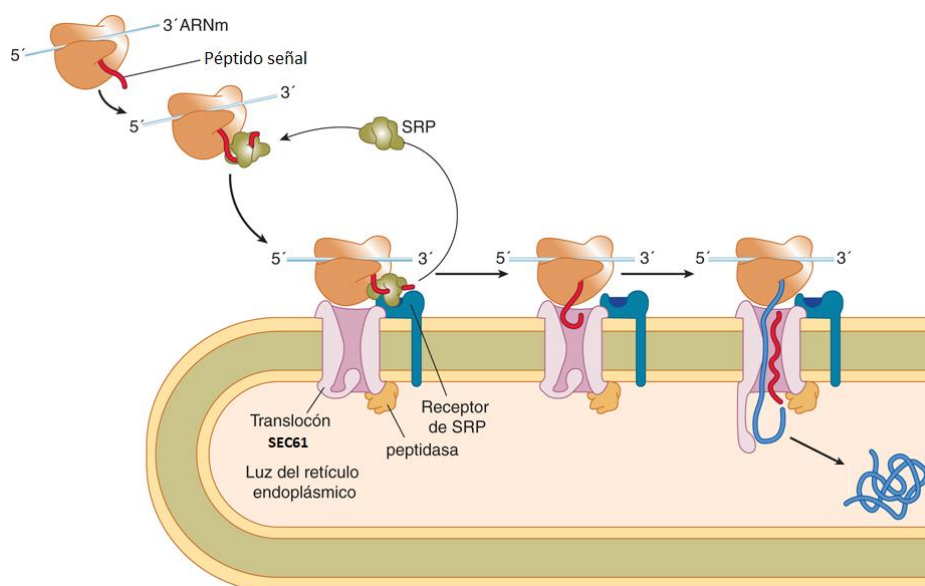


Figura 4. Translocación al RE de una proteína de vía secretora.

Durante la síntesis proteica pueden actuar sobre la cadena naciente diversas enzimas que la modifiquen. Dichas modificaciones, que tienen lugar antes de que se complete la síntesis, se conocen como co-traduccionales. Es posible que una modificación requiera de un determinado tiempo óptimo de traducción. Con lo cual se puede pensar que algunas modificaciones, como la glicosilación, podrían requerir de una determinada tasa de traducción en el momento en que la misma sucede. Así, la secuencia nucleotídica que codifica para la proteína y, no solo su secuencia aminoacídica, podrían determinar la eficiencia de la modificación.

Las glicoproteínas son resultado de la unión covalente de cadenas hidrocarbonadas a proteínas. La glicosilación es una de las modificaciones más importantes llevadas a cabo en el lumen del RE y en el complejo de Golgi. Muchos de los oligosacáridos que forman parte de glicoproteínas, están unidos a residuos de asparagina (Asn) mediante un enlace β -N-glicosídico. Cerca de un cuarto de las proteínas eucarionte son N-glicosiladas al ingresar a la vía secretoria. Los N-oligosacáridos que se transfieren a las cadenas polipeptídicas nacientes son homogéneos, relativamente simples y comunes a casi todas las células eucariotas. Luego de la transferencia, los N-glicanos son procesados en el RE manteniendo un espectro bastante uniforme de glicoformas. Durante el tránsito posterior por el complejo de Golgi los N-glicanos son sometidos a marcadas modificaciones en su estructura, generando una enorme diversidad química⁴⁸.

En el lumen del RE, un oligosacárido de alta manosa ($\text{GlcNAc}_2\text{Man}_9\text{Glc}_3$) es inicialmente transferido *en bloc* a partir del dolicol pirofosfato a una asparagina (Asn) presente en un secuón⁴⁹. Un secuón es una secuencia consenso (N-X-S/T, siendo $X \neq \text{Pro}$), necesaria pero no

suficiente para que se dé la *N*-glicosilación. La enzima que cataliza la transferencia del oligosacárido se denomina oligosacariltransferasa (OST), un complejo asociado a la membrana del RE, que está compuesto por siete subunidades en células humanas⁵⁰ (figura 5).

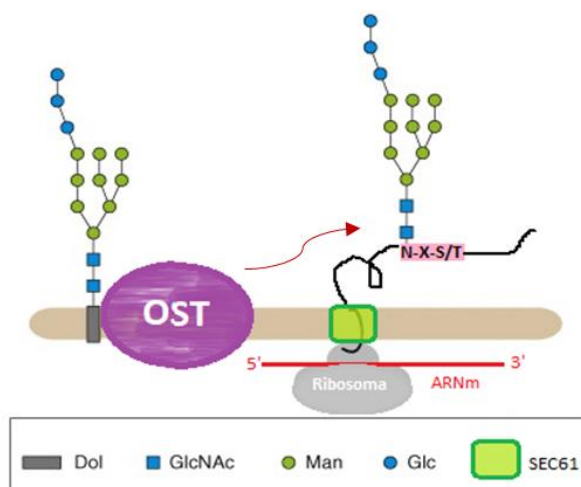


Figura 5. Esquema del proceso de *N*-glicosilación dentro del lumen del RE.

La subunidad de la OST encargada de la catálisis enzimática se denomina STT3. La misma reconoce ambos sustratos, al lípido dolicol oligosacárido dador y a la secuencia proteica aceptora. En algunos organismos como *T. cruzi* la STT3 es la única subunidad presente. En años recientes el grupo del Dr. Reid Gilmore demostró que en eucariotas superiores existen dos isoformas de esta subunidad, denominadas STT3A y STT3B⁵¹. La STT3A se encontraría en complejos de OST principalmente asociados al translocón Sec61 y su función sería la de glicosilar a los secuencias a medida que emergen en el lumen del RE. Por otra parte, los complejos que presentan la STT3B focalizarían su actividad sobre aquellos secuencias que quedaron vacantes luego de su ingreso al RE, siempre y cuando los mismos se encuentren accesibles al sitio activo de la enzima. Esta duplicación del gen de la subunidad catalítica permite a las células tener dos complejos OST actuando secuencialmente, maximizando entonces la eficiencia de la *N*-glicosilación. En particular, se pudo observar que la STT3B tiene un rol importante en la ocupación de secuencias cercanos al extremo C-terminal de las proteínas ó al péptido señal^{52,53}. En dicho escenario, la exposición al medio de un secuencia una vez que la proteína se pliega pasa a ser una variable relevante para su ocupación, así como también la cinética y estabilidad del plegado.

La glicosilación resulta importante para el correcto plegado de muchas proteínas, puesto que los *N*-glicanos facilitan la interacción con chaperonas del RE como calnexina y calreticulina. Además, la modificación aumenta la estabilidad proteica, protege de la agregación enmascarando superficies hidrofóbicas expuestas y del ataque por proteasas. También, modula

el direccionamiento a sitios sub y extracelulares, así como las interacciones célula-matriz y célula-célula.

Por otra parte, en virtud de su volumen y naturaleza hidrofílica, los *N*-glicanos alteran las propiedades biofísicas de los polipéptidos nacientes, aumentando su solubilidad y restringiendo la conformación local del polipéptido⁵⁴.

La presencia de un secuón no garantiza que la *N*-glicosilación suceda. Dado que existen diversos factores que afectan la eficiencia de este proceso, la ocupación de estos sitios es variable, dando origen a un fenómeno conocido como macroheterogeneidad de glicoformas (variabilidad en el grado de ocupación de los sitios).

La eficiencia de glicosilación depende de la disponibilidad de precursores unidos al lípido dolicol, de la actividad de la OST y de la estructura primaria, secundaria y terciaria de la proteína⁵⁵. El grado de ocupación de los sitios potenciales de *N*-glicosilación depende de la secuencia proteica primaria, particularmente de los aminoácidos que constituyen al secuón y de los que rodean a la secuencia consenso. Los aminoácidos vecinos a una Asn glicosilable influyen tanto las interacciones específicas con la OST como el plegado local, estabilidad y dinámica de las proteínas. En particular, es bien conocido que los secuones N-X-T son mucho más eficaces que los N-X-S^{56,57} debido a una mayor afinidad por el sitio activo de la OST. Además, la posición +1 (residuo "X") respecto de la Asn posee un marcado efecto en la glicosilación: mientras que los aminoácidos hidrofóbicos, voluminosos o ácidos reducen la ocupación de secuones, aquellos residuos pequeños, hidrofílicos ó básicos resultan en secuones altamente modificados. Una posible explicación a esto podría ser que la carga del aminoácido "X" influiría en la capacidad de la OST de unir simultáneamente al secuón y al precursor dolicol-PP-oligosacárido cargado negativamente⁵⁸.

Dado que la *N*-glicosilación es un proceso co-traducciona, se cree que la eficiencia del mismo resulta de una competencia entre la velocidad de plegado de la proteína y la adición del oligosacárido precursor unido al lípido dolicol⁵⁹.

En sus investigaciones, Malkin y Rich⁶⁰ (año 1967) y, Blobel y Sabatini^{61, 62} (año 1970) estimaron que dentro de la subunidad mayor del ribosoma caben entre 35 a 39 residuos del extremo carboxi terminal de la cadena naciente y, que se requiere un mínimo de 20 residuos para atravesar la bicapa lipídica. Esas consideraciones sugieren que antes de que un sitio

particular pueda estar accesible al lumen del RE una longitud mínima de 55 a 60 residuos debe ser sintetizada. Estudios posteriores determinaron que 65 residuos son necesarios para cubrir la distancia entre el sitio P del ribosoma y el sitio activo de la OST (**figura 6**). Dicha longitud aumenta a 71 residuos cuando se encuentra un stop transfer de 18 residuos hidrofóbicos de extensión en la cadena: mientras que cadenas nacientes no hidrofóbicas podrían adoptar una conformación totalmente extendida durante el pasaje a través del complejo de translocación, las secuencias hidrofóbicas parecen formar estructuras más compactas (posiblemente hélices) dentro del translocón⁶³.

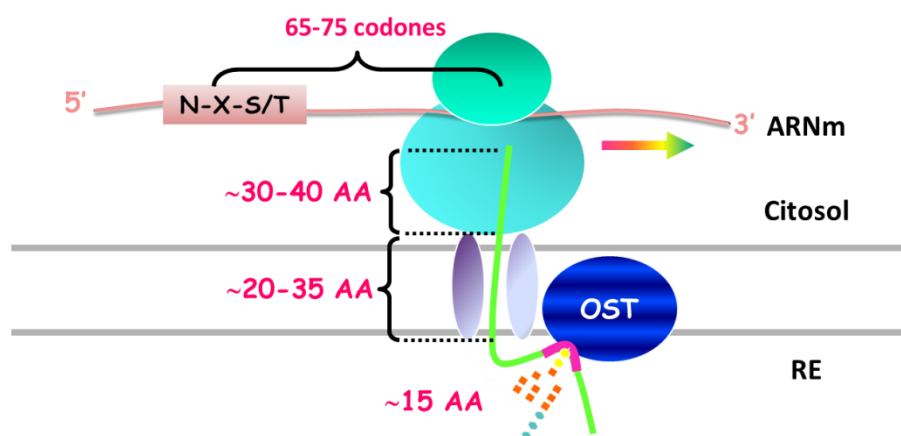


Figura 6. Se muestran las “distancias” ocupadas por aminoácidos dentro del ribosoma y del translocón.

Cualquier proceso que ocurra co-traduccionalmente es pasible de ser modulado por la velocidad de traducción. Un ejemplo paradigmático es la expresión de variantes sinónimas de la proteína de multiresistencia a drogas (MDR1)⁶⁴. Esta proteína es una bomba localizada en la membrana plasmática que expulsa del citosol compuestos orgánicos de estructura diversa. Su función biológica está relacionada a mecanismos de detoxificación celular y es uno de los principales responsables de la resistencia a la quimioterapia durante el tratamiento del cáncer. Se han encontrado mutaciones en el gen de MDR1 que aumentan su actividad de bombeo y las personas que portan dichas mutaciones desarrollan una mayor resistencia a las drogas quimioterápicas. Llamativamente, las mutaciones que presentan estos pacientes son sinónimas. Si bien estas no alteran la estructura primaria de la proteína, sí afectan a su estructura terciaria, aumentando su actividad como transportador. Más recientemente, se encontró que en diversas proteínas multidominio las regiones que conectan los dominios están enriquecidas en codones poco abundantes^{65, 66, 67}. Se pudo observar que cuando el ribosoma traduce dichas proteínas realiza una pausa en la zona interdominios, aumentando la eficiencia del proceso de plegado. Se ha propuesto que dicha pausa permite que comience el plegado del primer dominio traducido

apenas emerge del ribosoma, separando temporalmente su plegado del dominio siguiente en la secuencia. Si bien la aparición de codones lentos puede disminuir el rendimiento en cuanto a los niveles de expresión, esto se compensa con un aumento en la eficiencia del plegado.

Estos conocimientos cambiarán el modo de optimizar la expresión heteróloga mediante la armonización de codones ya que, en algunos casos, el resultado óptimo desde el punto de vista de los niveles de expresión y la eficiencia de plegado se obtendría incorporando pausas en regiones estratégicas.

Por otra parte, se ha visto que la velocidad de traducción puede influir sobre las modificaciones químicas de las proteínas. Un claro ejemplo es la degradación de las isoformas beta y gamma de la actina⁶⁸. Ambas proteínas pueden ser arginiladas en su extremo amino terminal al salir del ribosoma. Mientras que la beta-actina arginilada es estable, la gamma-actina arginilada es modificada por ubiquitina y degradada inmediatamente. Se demostró que la diferencia entre ambas isoformas radica en su velocidad de traducción. Una pausa local durante la traducción de la isoforma gamma, previa a su plegado, permite la exposición temporal de la Lys18, residuo que normalmente queda oculto en la estructura terciaria. La exposición de la Lys18 permite su modificación por ubiquitina, marca que conduce a la proteína a degradación. La isoforma beta, al no presentar dicha pausa, se traduce rápidamente y la proteína se pliega ocultando a la Lys18 que no resulta modificada.

Dado que la velocidad de traducción puede alterar el resultado de los procesos co-traduccionales, sería relevante explorar si la *N*-glicosilación es uno de estos procesos.

El pasaje del polipéptido nascente por el sitio activo de la OST está limitado a la tasa de elongación (6 aminoácidos/segundo en mamíferos). ¿Existe alguna evidencia de que la velocidad de traducción podría estar modulando la *N*-glicosilación? La evidencia preliminar sugiere que este sería el caso. La misma se basa en un trabajo donde se ha observado que cuando un secuón es movido hacia el C-terminal de la proteína, se alcanza un punto donde el mismo deja de ser ocupado. Los secuones ubicados a menos de 55 residuos del extremo C-terminal de las proteínas se ocupan pobremente. Una posible explicación sería que cuando el ribosoma alcanza al codón STOP en el ARNm, la cadena polipeptídica se soltaría rápidamente y eso no le daría tiempo a la OST para realizar la catálisis de la *N*-glicosilación⁶⁹. En este sentido, estudios estadísticos han mostrado que la frecuencia de tripéptidos Asn-X-Ser/Thr no glicosilados aumenta a medida que los mismos se acercan al extremo C-terminal de la proteína⁷⁰.

En resumen, la evidencia experimental acerca de los diversos procesos regulados por la velocidad de traducción y el conocimiento actual acerca del proceso de *N*-glicosilación, apoyan a un modelo en donde ambos procesos podrían estar íntimamente relacionados.

El mercado de proteínas terapéuticas *N*-glicosiladas representa ventas multibillonarias y crece más de un 10% cada año. Eso requiere de plataformas de producción rentables que conlleven a una *N*-glicosilación correcta y homogénea. Los productos farmacéuticos poseen controles estrictos en cuanto a la homogeneidad del producto. Como fue mencionado, la ocurrencia de un secuón no implica su modificación con un *N*-glicano. Existen numerosos ejemplos de secuones que no son ocupados, o que lo son parcialmente⁷¹. Esta macroheterogeneidad de las glicoproteínas implica una variabilidad en el grado de ocupación de los sitios de *N*-glicosilación, siendo uno de los principales problemas a solucionar por la industria farmacéutica que produce proteínas de interés sanitario. Muchas de estas proteínas son *N*-glicosiladas (interferón gamma, eritropoyetina, anticuerpos monoclonales, factor tisular, etc) y usualmente se expresan en líneas celulares animales cultivadas en fermentadores. Los organismos estatales de regulación requieren que las proteínas fármaco presenten un determinado perfil de *N*-glicosilación, siendo la macroheterogeneidad uno de los principales escollos para cumplir con dichas exigencias. Por este motivo, existe un enorme interés en optimizar la ocupación de los secuones de *N*-glicosilación. A pesar del enorme conocimiento acumulado en el área, aún no está claro por qué ciertos secuones son poco eficientes.

Un buen ejemplo de macroheterogeneidad lo constituye el interferón beta humano, proteína que contiene un sitio de *N*-glicosilación conservado en la Asn80 y que es comercializada para el tratamiento de esclerosis múltiple. Comparando IFN- β comercializados se encontró que la actividad antiviral de la forma glicosilada es diez veces mayor respecto del producto generado en *E. coli* (que no *N*-glicosila). En cultivos de células CHO se observa que la Asn80 no siempre es glicosilada, existiendo entonces macroheterogeneidad⁷².

Resulta evidente la importancia que tendría poder controlar la macroheterogeneidad de los bioterapéuticos.

HIPÓTESIS Y OBJETIVOS

VII. Hipótesis de trabajo.

La distancia que debe atravesar un polipéptido desde el sitio P del ribosoma hasta el sitio activo de la oligosacariltransferasa ha sido determinada experimentalmente entre unos 65-75 residuos. Esto quiere decir que cuando un secunón llega a la OST, el ribosoma está leyendo al ARNm en una posición que está ~65 codones río abajo hacia el extremo 3' (**figura 7**).

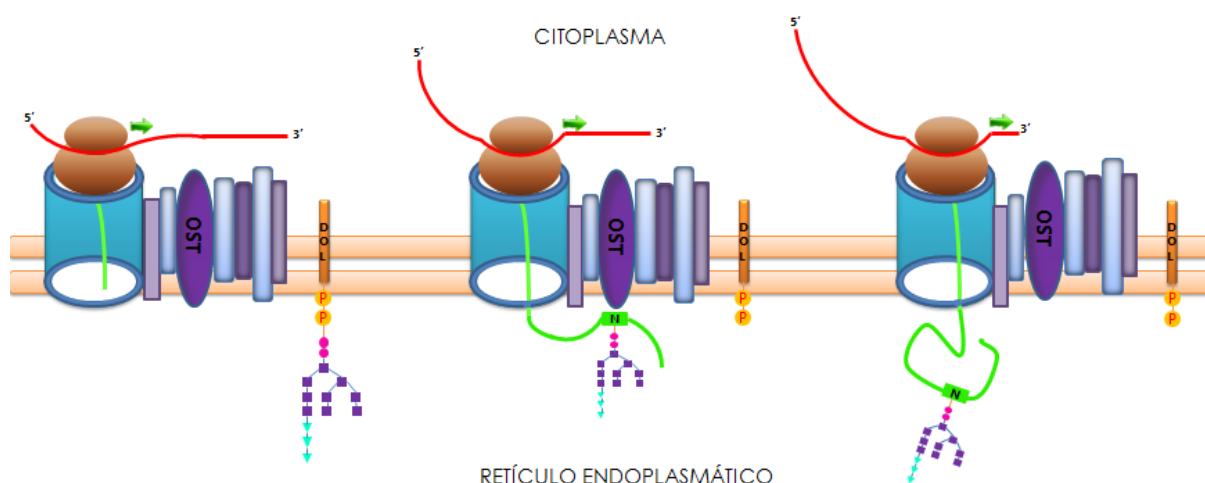


Figura 7. Esquema donde puede observarse que desde el momento en que emerge del ribosoma una asparagina (N) de un secunón, hasta que alcanza el sitio activo de la OST, el ribosoma ha avanzado hacia el extremo 3' del ARNm.

La velocidad de traducción de esta región del ARNm podría influir en el tiempo de residencia del secunón en el sitio activo de la OST, y por ende podría modular su grado de ocupación. Hipotetizamos que aquellos ARNm que presentasen una velocidad de lectura menor (pausas) en dicha posición mostrarían una eficiencia mayor en la ocupación de los secunones. Un dato que apoyaba a esta hipótesis es que generalmente los secunones que se encuentran a menos de 55 residuos del extremo C-terminal de las proteínas son poco utilizados^{69,73}. En esos casos se especula que la lectura del codón STOP libera a la proteína naciente del ribosoma, provocando un tránsito rápido por el sitio activo de la OST, disminuyendo así la eficiencia de *N*-glicosilación de los secunones presentes en dicha región.

Nuestra hipótesis consistió en que la presencia de pausas durante la traducción aumentaría la eficiencia de *N*-glicosilación de proteínas.

Según el modelo planteado, las pausas en la traducción incrementarían la eficiencia de *N*-glicosilación siempre y cuando éstas ocurrieran en la distancia correcta respecto del secunón

dentro del ARNm (**ver figura 8**). Se esperaba que las construcciones donde la pausa no ocurriera en la localización adecuada, no vieran afectada la eficiencia de *N*-glicosilación.

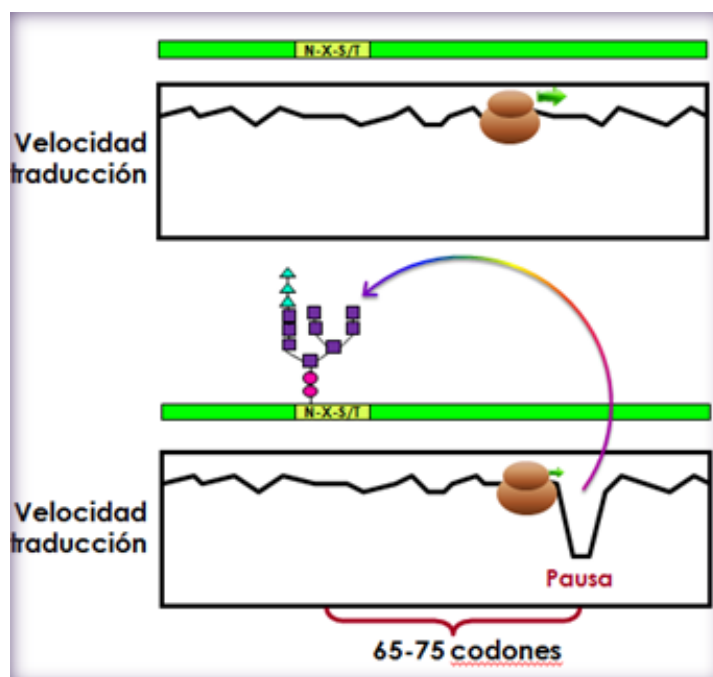


Figura 8. Esquema de la hipótesis de trabajo, en el que se observa que una disminución en la velocidad de síntesis proteica permitiría la ocupación de un sitio de *N*-glicosilación, cuando la misma ocurriera a la distancia en la que el secuéon resultase accesible a la OST.

Como segunda hipótesis de trabajo se plantea que alteraciones en la velocidad de plegado y/o en la estabilidad termodinámica de las proteínas podrían modular la eficiencia de ocupación de secúones. Un mayor tiempo de exposición al medio de los secúones en proteínas parcialmente plegadas y/o una mayor labilidad estructural permitiría una mejor ocupación de aquellos sitios que quedaron vacantes durante el ingreso. Dado que este efecto se vería principalmente sobre los procesos que ocurren luego de la translocación de las proteínas al lumen del RE, los mismos estarían mediados principalmente por la subunidad STT3B.

VIII. Objetivos de trabajo.

El objetivo general de la primera parte del proyecto consistió en explorar la relación entre la velocidad de traducción y la eficiencia de la *N*-glicosilación de proteínas. Específicamente, se estudió el efecto del empleo de codones sinónimos sobre la eficiencia de *N*-glicosilación de una variante de la proteína verde fluorescente (*cf-sfGFP*) diseñada especialmente para poder evaluar la hipótesis del proyecto.

Las pausas mencionadas serían generadas por la presencia de codones poco frecuentes en el extremo 3' del ARNm, tal que la lectura del mensajero por parte del ribosoma en dicha zona resultase enlentecida. Los codones poco frecuentes fueron colocados en una posición en la cual se esperaba que el secuón hubiera alcanzado el sitio activo de la OST.

Los objetivos específicos fueron:

1) Diseñar en base a bibliografía y herramientas de modelado molecular una GFP que permita poner a prueba la hipótesis de trabajo. La misma deberá presentar dos variantes: una proveniente de un ARNm de lectura rápida y otra, cuyo ARNm poseerá en sus últimos 63 codones a aquellos que permitan una traducción local lenta.

2) Realizar el clonado en vectores de expresión en mamíferos que permitan el direccionamiento a retículo endoplásmico, citoplasma y medio extracelular.

3) Introducir por mutagénesis sitios de *N*-glicosilación en diferentes posiciones de la secuencia proteica para así evaluar la hipótesis de trabajo.

4) Transfectar las variantes de *N*-glicosilación en cultivos celulares de líneas humanas y cuantificar la relación *proteína glicosilada/proteína total*.

5) De determinarse que la velocidad de traducción modifica el grado de *N*-glicosilación:

i) se procederá a acotar la distancia en que debe disminuir la velocidad de síntesis proteica para que aumente la *N*-glicosilación, acortando la extensión del segmento de lectura lenta del ARNm.

ii) se aplicará este conocimiento para optimizar la ocupación de los sitios de *N*-glicosilación de diversas proteínas de interés biotecnológico.

En una segunda parte del trabajo, en base a los resultados obtenidos, nos propusimos estudiar la relación entre la estabilidad termodinámica y la eficiencia de *N*-glicosilación de proteínas. Planteamos que una menor estabilidad termodinámica, permitiría una mayor ocupación de los secuones debido a fluctuaciones en la estructura que darían mejores chances a la oligosacariltransferasa para transferir el *N*-glicano. Para esto, se construyeron variantes de la sfGFP de estabilidad alterada y se estudió la ocupación co- y post-traducciona de los secuones.

MATERIALES Y MÉTODOS

IX. ADN, clonado y mutaciones.

La síntesis de la secuencia codificante para la *cf-sfGFP* rápida, no glicosilada, así como el segmento lento fue encargado a GenScript.

Rápida o fast	ATGAGCAAAGGTGAAGAACTATTTACGGGTGTAGTACCGATACTAGTAGAACTAGATGGTGATGTAA ATGGTCATAAATTTTCGGTACGTGGCGAGGGCGAGGGCGACGCTACAAACGGCAAGCTGACACTGAA GTTCATTTCTACAACAGGCAAGCTGCCAGTGCCATGGCCAACACTGGTGACAACACTGACATACGGCG TGCAGATGTTCTCTAGATACCCAGACCACATGAAGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAG GGCTACGTGCAGGAGAGAACAATTTCTTTCAAGGACGACGGCACATACAAGACAAGAGCTGAGGTGA AGTTCGAGGGCGACACACTGGTGAACAGAATTGAGCTGAAGGGCATTGACTTCAAGGAGGACGGCA ACATTCTGGGCCACAAGCTGGAGTACAACCTCAACTCTCACAACGTGTACATTACAGCTGACAAGCAG AAGAACGGAATTAAGGCTAACTTCAAGATTAGACACAACGTGGAGGAC GGATCC GTGCAGCTGGCTG <u>ACCACTACCAGCAGAACACACCTATCGGCGACGGCCCTGTGCTGCTGCCTGACAACCACTACCTGTCT</u> <u>ACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGAGAGACCACATGGTGCTGCTGGAGTTCGTGA</u> <u>CCGCTGCTGGCATCACACGGCATGGACGAGCTGTACAAG</u>
Segmento lento o slow	GGATCC GTACAACACTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCCGGTACTACTACC <u>GGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACGTGATCATATGG</u> <u>TACTACTAGAATTGTAAACGGCGGGGTATAACGCATGGTATGGATGAACTATATAA</u>

Tabla 2. Secuencias encargadas a GenScript. En negrita se indica el sitio de corte para la enzima BamHI y se muestra subrayado el segmento modulador de la velocidad.

Se liberó mediante enzimas de restricción la secuencia de la *cf-sfGFP* rápida y se la subclonó en un pcDNA3.1 que contenía un codón de inicio de traducción, seguido del péptido señal de la prolactina y, en el otro extremo, un epítipo 3xHA (3x YPYDVPDYA) seguido de un codón stop.

Una vez obtenida la construcción rápida, se liberó el segmento que modula la velocidad y se lo ligó con el segmento lento encargado.

Las digestiones con endonucleasas de restricción se realizaron respetando los medios y condiciones recomendadas por los proveedores de cada enzima en particular. En todos los casos la reacción se dejó transcurrir overnight a 37°C, empleándose de 1 µl de enzima por tubo. Se inactivó por calor a aquellas enzimas que era posible.

Las construcciones o fragmentos ADN fueron purificados a partir de geles de agarosa utilizando un kit de purificación (QIAquick® de QIAGEN) siguiendo las instrucciones del fabricante.

Para las reacciones de ligación se utilizaron 10 µl de inserto + 1 µl de vector, en presencia de 1 µl de una dilución (1:20) de la enzima T4 DNA Ligasa y 2 µl de buffer ligasa, incubando toda la noche en una solución de hielo-agua. Las ligaciones fueron dializadas durante 30 min contra

agua miliQ a través de un un filtro Millipore. Luego, el ADN recuperado fue electroporado en bacterias *E. coli* cepa DH10b.

X. Cepa bacteriana y medio de cultivo empleado.

Para los trabajos de clonado y mutagénesis se usaron bacterias DH10b electrocompetentes.

El medio LB (Luria-Bertani), compuesto de 10 g/l tripteína bacteriológica, 5 g/l extracto de levadura y 10 g/l NaCl; y el LB-agar, medio LB con 2 % p/v agar, se utilizaron para el crecimiento bacteriano. Para seleccionar clones bacterianos, se suplementó el medio con 200 mg/l de ampicilina.

XI. Transformación bacteriana

Se descongelaron 200 µl de células competentes y se mezclaron con 10 µl del producto de ligación. Se utilizó el equipo de electroporación de BIO-RAD Gene Pulser®, fijando los parámetros en 2,5 kV y 200 Ohms. Luego de la electroporación, se agregó a la cubeta 700 µl de medio LB y se incubó durante 1 h a 37°C con agitación fuerte. Diferentes volúmenes de las células transformadas fueron sembrados en placas de Petri con LB agar y ampicilina, y se lo dejó crecer en estufa a 37°C overnight.

Al día siguiente, se estriaron clones aislados en LB agar con ampicilina y se dejó crecer a 37°C. Con el mismo tip con que se estrió el clon, se largó un cultivo de LB ampicilina que se dejó en agitación a 37°C.

XII. Extracción plasmídica

Se obtuvo el ADN plasmídico de los cultivos largados overnight mediante la técnica de lisis alcalina (miniprep) como se detalla a continuación:

- a) Centrifugar los cultivos en eppendorf de 1,5 ml y descartar sobrenadante.
- b) Agregar 200 µl de solución P1 (50 mM Tris/HCl, 10 mM EDTA, pH= 8.0) y resuspender en vórtex.
- c) Poner 300 µl de solución P2 (0,2 M NaOH, 1% SDS) y mezclar el tubo por inversión muy suavemente. Incubar en hielo 5 minutos.
- d) Agregar 300 µl de solución P3 (3 M KAc, pH= 4.8) y mezclar el tubo por inversión muy suavemente. Incubar en hielo 5 minutos
- e) Centrifugar en microcentrífuga a máxima velocidad durante 15 minutos.

f) Tomar el sobrenadante y pasarlo a otro eppendorf conteniendo 2 μ L de RNAsa A 10 mg/ml. Incubar a 37°C durante 30 minutos.

g) Hacer dos extracciones con 400 μ l cloroformo, mezclando bien y centrifugando 5 minutos a velocidad máxima.

h) Pasar al sobrenadante a un eppendorf nuevo con 700 μ l de isopropanol, agitar suavemente e incubar en hielo durante 2,5h.

h) Centrifugar en microcentrífuga refrigerada a máxima velocidad durante 25 minutos.

i) Descartar el sobrenadante y lavar con 500 μ l de etanol 70%. Descartar el etanol. Dejar evaporar el resto de etanol con el eppendorf abierto.

j) Poner 32 μ l de agua miliQ y resuspender precipitado. Agregar 48 μ l de una solución de PEG8000 13 % p/v, NaCl 4M, agitar e incubar durante 40 minutos en hielo.

k) Centrifugar a velocidad máxima en centrifuga refrigerada durante 25 minutos.

l) Retirar sobrenadante con tip y lavar con 400 μ l de etanol 70%.

m) Centrifugar durante 5 minutos, retirar sobrenadante y resuspender en 35-50 μ l de agua miliQ.

Se chequeó la obtención de ADN de tamaño esperado por electroforesis en agarosa y se mandaron a secuenciar clones al servicio brindado por el Instituto Leloir a cargo de la Lic. Marta Bravo, o a la empresa Macrogen (Korea).

XIII. Generación de secuencias

Obtenidas las variantes no glicosiladas rápida y lenta de la *cf-sfGFP*, se dio inicio a la generación individual de las mutantes de *N*-glicosilación. Para esto se empleó en un comienzo, la mutagénesis con PCR por extensión solapada (**figura 9**). El método de extensión solapada requiere de 2 cebadores que contienen la mutación deseada (forward y reverse) y otros 2 que se unen a los extremos del producto final.

La técnica se lleva a cabo en dos etapas. En la primera, se realizan dos reacciones de PCR por separado utilizando oligonucleótidos externos, que hibridan en ambos extremos de la secuencia de ADN clonada en el vector pcDNA3.1, y oligonucleótidos internos, que contienen la mutación puntual. Los productos de PCR obtenidos en ambas reacciones se purifican a partir de un gel de agarosa y se emplean como molde para la segunda etapa, en la que se realiza una única reacción de PCR. Los diez primeros ciclos se hacen sin el agregado de cebadores para permitir la hibridación y elongación del producto completo. Finalizados esos diez ciclos, se

agrega a la mezcla de reacción 5 µl de cada uno de los oligonucleótidos externos y se continúa con el mismo programa durante 36 ciclos.

Los oligonucleótidos externos contienen los sitios de restricción para las enzimas EcoRI y XhoI, respectivamente, para realizar el clonado posterior de las mutantes de *N*-glicosilación en el vector pcDNA3.1.

El producto de PCR obtenido se analizó mediante electroforesis en gel de agarosa 2%, desarrollando durante 30 minutos a 100 V, purificándose luego la banda buscada.

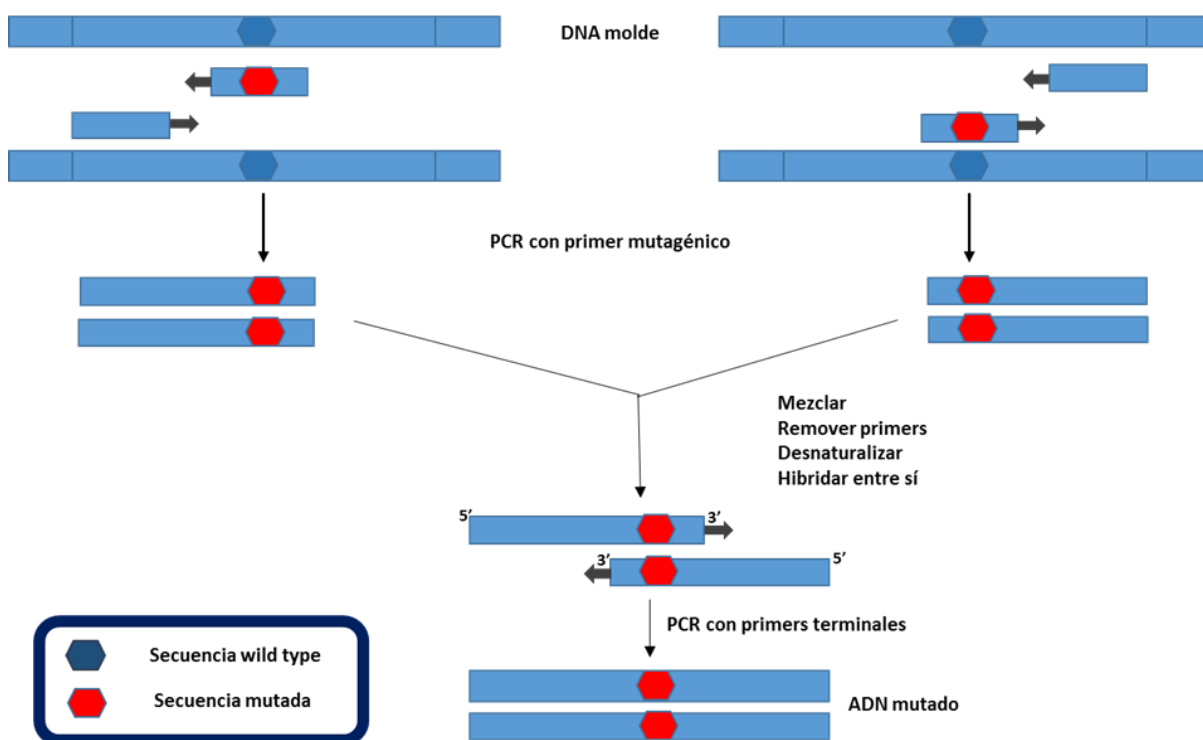
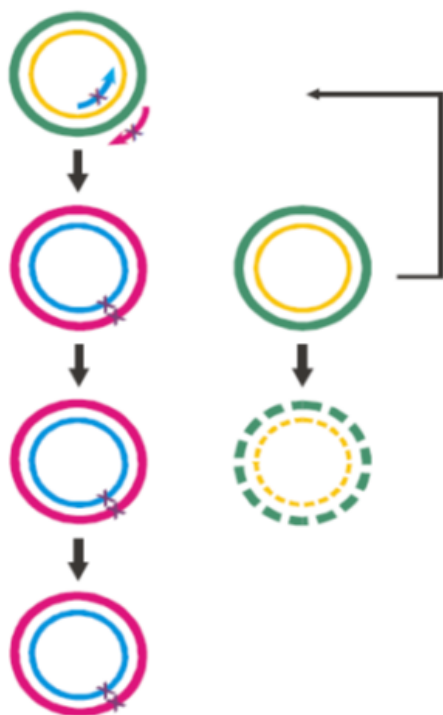


Figura 9. Esquema del proceso de PCR. Primero, se amplifica un fragmento 5' y otro 3', que portan la mutación y solapan entre sí. Estos productos se usan como cebadores en otra reacción para producir el ADN mutado de longitud completa. Finalmente se completa el proceso, realizando unos ciclos con los cebadores de los extremos.

A partir de mediados del segundo año de tesis se usó la técnica de mutagénesis dirigida de Agilent (QuickChange) que emplea a la enzima pfu. Los cebadores a emplear fueron diseñados en la web <http://www.genomics.agilent.com/primerDesignProgram.jsp> y el protocolo puede verse en la **figura 10**.



Síntesis de la hebra mutante

Realizar ciclos térmicos para:

- 1) Desnaturalizar el ADN molde
- 2) Hibridar los primers con la mutación deseada
- 3) Extender los primers con Pfu ADN polimerasa

Digestión del molde con DpnI

Digiere el ADN parental metilado y hemimetilado con DpnI

Transformación

Transformar la molécula mutada en células competentes para la reparación del nick

Figura 10. Esquema de mutagénesis modificado a partir de <http://www.agilent.com/cs/library/usermanuals/Public/200523.pdf>

Para 50 µl de mezcla de reacción se ponen 0,5 µl de BSA [100X], 5 µl de buffer pfu [10X], 1 µl de MgSO₄ [50mM], 1 µl de dNTPs [10mM] c/u, 1 µl de pfu, 4 µl de ADN molde [1:30], 5 µl de cada cebador y agua mQ hasta completar el volumen final.

El programa de PCR utilizado fue:

1)	95 °C x 2'
2)	95 °C x 30"
3)	50 °C x 30'
4)	73 °C x 13' (2' x kpb)
6)	73 °C x 5'

5) 25 veces

XIV. Líneas celulares utilizadas.

- COS-7
- HeLa
- HEK 293 silvestre
- HEK 293 STT3A^{-/-}
- HEK 293 STT3B^{-/-}

Las tres líneas HEK 293 fueron cedidas gentilmente por el Dr. Reid Gilmore (Universidad de Massachussets, EE.UU). Las líneas KO fueron generadas en su laboratorio mediante la técnica de edición genética CRISPR.

XV. Medios de cultivo.

El medio DMEM alta glucosa de Gibco se utilizó para el crecimiento y mantenimiento de las células. La esterilización de los medios de cultivo se realizó por filtración a través de filtros estériles de Millipore, cuyo tamaño de poro fue de 0,2 μm . Se utilizó suero fetal bovino (SFB) de Natocor al 10% v/v para suplementar el medio de cultivo.

XVI. Revitalización de líneas celulares.

Para descongelar líneas celulares criopreservadas, el criotubo fue retirado del tanque de N_2 líquido y sumergido rápidamente en un baño de agua a 37°C. La suspensión celular descongelada fue plaqueada en una P100 con 8 ml de DMEM con SFB. A primera hora de la mañana siguiente, se realizó un lavado con PBS y se colocó medio fresco para eliminar el DMSO.

XVII. Mantenimiento de líneas celulares.

Las líneas celulares empleadas se cultivaron en estufa gaseada a 37°C, en atmósfera de CO_2 al 5 %, saturada de humedad.

XVIII. Criopreservación de líneas celulares.

Cultivos celulares en fase de crecimiento exponencial fueron tripsinizados. Se procedió a contar las células en cámara de Neubauer y luego se centrifugó durante 5 minutos a 800 rpm. El sobrenadante fue descartado y el pellet celular resuspendido en una solución de SFB y DMSO 10% (v/v) tal que se congelaran por criovial entre $1,2\text{-}2 \times 10^6$ células por ml. La suspensión celular fue enfriada gradualmente: a 4°C durante 20 minutos, a -80°C durante la noche y al día siguiente, los criotubos fueron conservados a -196°C en tanque de N_2 líquido.

XIX. Transfección.

Se sembraron células en una placa de pocillos múltiples 24 hs antes de la transfección. En el caso de las placas de 24 pocillos fueron sembradas 65000 células/pocillo y, en las placas de 12 pocillos, 120000 células/ pocillo.

Al día siguiente de la siembra, se cuantificó la concentración de ADN con un Nanodrop, para transfectar 1 µg de ADN/pocillo en los MW24 y 2 µg ADN/pocillo en los MW12. Se observó que los pocillos estuvieran homogéneos y en confluencia adecuada.

Para las transfecciones se empleó el reactivo Lipofectamine™ 2000 de Invitrogen de acuerdo a las instrucciones del fabricante. Se realizó una solución de lipofectamina + DMEM sin suero y, por otro lado, una solución de ADN + DMEM sin suero. Se mezclaron ambas soluciones y se las incubó 10 minutos a temperatura ambiente. Se sacó la placa sembrada de la estufa, se quitó el medio con suero de cada pocillo y fue reemplazado por medio sin suero. Luego se agregó la solución de lipofectamina + ADN + medio sin suero a cada pocillo, se agitó en forma de cruz a la placa y se la dejó en la estufa de cultivo durante 2,5 h. En ese momento, se quitó el medio y se lo reemplazó por un volumen equivalente de DMEM con suero. Se incubó en estufa durante un tiempo determinado, dependiendo del experimento. Los volúmenes de medio y lipofectamina empleados fueron los indicados en la hoja de datos del reactivo según la placa multipocillo empleada.

XX. Lisis celular.

Al día siguiente de la transfección o transcurrido el tiempo indicado, se chequeó la expresión de la proteína verde fluorescente bajo microscopio de fluorescencia. Se quitó el medio de cada pocillo, se realizó un lavado con PBS 1X durante 2 min y se agregó un volumen de buffer lisis con inhibidores de proteasas. Se incubó durante 1 h en hielo sobre un agitador de vaivén.

El buffer lisis se compone de Tris-HCl 25mM pH=7.5, NaCl 150mM, Tritón X-100 1 % v/v, EDTA 1 mM. Antes de usar, se agregaron los inhibidores de proteasas E-64, leupeptina, pepstatina, y PMSF. Luego de la incubación en hielo, se empleó un tip para raspar el pocillo y finalmente, se pasó el lisado celular a un tubo eppendorf con cracking buffer 5X, β-MSH 10 % v/v. Se colocó en un termobloque a 100°C durante 10 min, se dejó enfriar y las muestras fueron alicuotadas y conservadas a -80°C hasta su uso.

XXI. Marcación isotópica celular.

A las células plaqueadas en una placa multipocillo se las transfectó con el ADN correspondiente a la versión lenta ó rápida de la *cf-sfGFP*. Transcurridas 24 hs, se procedió al hambreado de las células: primero se quitó de cada pocillo el DMEM y se lo lavó con PBS, luego se colocaron 400 µl de *Gibco*™ DMEM, sin glucosa, glutamina, metionina y cisteína complementado con L-glutamina [4 mM] y se dejó en estufa a 37°C durante 20 minutos. En el laboratorio bajo campana se fueron realizando las marcaciones en distintos tiempos: 0', 2', 4',

6', 8' con 400 µl de medio de hambreado con EasyTag™ EXPRESS³⁵S Protein Labeling Mix, [35S]-, 50 µCi de PerkinElmer, todo en un termoagitador a 37°C. A los 10' de la primera marcación, se comenzó por quitar el medio al último punto de marcado y se siguió en línea hasta llegar al primer punto. Se lavó con buffer PBS y se colocaron 400 µl de medio de caza compuesto por DMEM con SFB 10 % v/v y una concentración final de 2 mM de cisteína y metionina. Se lavó nuevamente con PBS y se incubó a la placa en hielo con 100 µl de buffer lisis por pocillo durante 1h.

XXII. Inmunoprecipitación.

En el caso de las inmunoprecipitaciones, se siguió el procedimiento de sembrado, transfección y lisis celular hasta obtener el lisado total por raspado con el tip. El lisado se transfirió a un eppendorf con bolitas de proteína G-agarosa lavadas, y se incubó en agitación de vaivén durante 2 hs para eliminar todo lo que pudiera pegarse inespecíficamente a las mismas. Se centrifugó a 1000 rpm durante 2 min y el sobrenadante fue transferido a un tubo eppendorf conteniendo proteína G-agarosa lavada + anticuerpo anti-HA. Esto se dejó incubando en cámara fría hasta el día siguiente con agitación de vaivén.

Al otro día, se realizó una centrifugación a 1000 rpm durante 2 min, se descartó el sobrenadante y se realizaron lavados con PBS 1X, NP40 0,5 % v/v, centrifugando a 1000 rpm durante 2 min y descartando el sobrenadante. Al pellet se lo resuspendió con 40 µL de PBS 1X en vórtex, se le agregó 10 µL de cracking buffer 5X, β-MSH 10 % v/v y se calentó a 100°C durante 10 min. En la corrida electroforética en gel de poliacrilamida, se sembró el sobrenadante de este último paso.

XXIII. Microscopía.

El microscopio invertido utilizado fue el Olympus CKX41, que posee una cámara Olympus Qcolor 5 acoplada y una lámpara de mercurio Olympus U-RFLT50. Las imágenes de fluorescencia fueron adquiridas utilizando el programa Qcapture Pro7, empleando el filtro de excitación azul (460-490 nm). Las mismas fueron tomadas directamente sobre la placa multipocillo en la cual se hubieran sembrado las células.

XXIV. Geles.

Se armaron geles de poliacrilamida 15 % y se sembró el total de la inmunoprecipitación. Luego de la electroforesis, el gel fue incubado unos minutos en glicerol 10 % y posteriormente secado ON en un secador de geles por vacío.

XXV. Revelado.

Se colocó al gel seco en contacto con una pantalla autoradiográfica dentro de un cassette fotográfico y se lo dejó exponiendo durante 15 días. Luego, se reveló utilizando un revelador Storm.

XXVI. Western Blot.

Electroforesis en gel de poliacrilamida. Las muestras guardadas a -80°C fueron descongeladas y sembradas en un gel 15 %, en un sistema de electroforesis standard de Hoefer. Se aplicó un voltaje de 90 V para el gel apilador y luego se corrió a 145 V hasta pasados 10 minutos desde el momento en que el frente de corrida salió del gel separador.

Transferencia. La misma se realizó durante la noche a 4°C en un sistema Mini Trans-Blot[®] Cell de BIO-RAD a 180 mA. La membrana utilizada fue de PVDF.

Bloqueo. La membrana de PVDF transferida fue bloqueada durante 1,5 h en una solución de leche 5 % v/v, TBS 1X.

Detección. Se empleó el anticuerpo anti-HA acoplado a peroxidasa (anti-HA-HRP) en una dilución (1:10.000 v/v) en TBS 1X y se incubó durante 1,5 h a la membrana a 20°C con agitación sobre un agitador térmico. Se recuperó la solución de anticuerpo y se realizaron lavados sucesivos con TBS1X durante 5 min, 3 min, 2 min, 5 min.

Para la reacción de quimioluminiscencia se empleó el kit ECL[™] Western Blotting Analysis System (GE Healthcare). Se incubó a la membrana durante 5 min a temperatura ambiente con 400 μL de solución 1 y 400 μL de solución 2 del kit de detección. Se retiró el exceso de reactivo y se colocó a la membrana dentro de un Nylon. Finalmente se utilizó el equipo LAS4000 para revelar y obtener las imágenes.

Cuantificación. Las bandas proteicas de las imágenes obtenidas fueron cuantificadas empleando el programa ImageJ.

Anticuerpos utilizados.

I) Anticuerpo monoclonal hecho en ratón, isotipo IgG1, contra el HA de Pierce[™] (2-2.2.14), número de catálogo #26183.

II) Anticuerpo monoclonal hecho en ratón, isotipo IgG1, contra el HA de Invitrogen (2-2.2.14) conjugado a HRP, número de catálogo #26183-HRP.

XXVII. Clonado y expresión de proteínas en *E. coli*

Las secuencias que codifican para *cf-sfGFP* ("M70") y *sfGFP* C70 ("C70") fueron amplificadas a partir de los vectores pcDNA3.1 empleados para su expresión en células de mamífero. La amplificación se realizó con la DNA polimerasa Pfu provista por el instituto Leloir, utilizando los cebadores JJC512 (5'-AATGCACCCCATATGAGCAAAGGTG-3') y JJC513 (5'-CGAGTTCTCGAGCTTGACAGC-3'). Se realizaron 40 ciclos de amplificación y una temperatura de apareamiento de 46°C. El producto de amplificación fue purificado mediante electroforesis en agarosa 2% y posterior elución en H₂O mQ. El inserto fue digerido ON a 37°C con 1 µl de cada una de las enzimas de restricción NdeI (**CA-TATG**, NEB cat. # R0101S) y XhoI (**C-TCGAG**, NEB cat. # R0146S). El producto fue calentado a 65°C durante 20 min y purificado empleando columnas QIAquick, eluyéndolo en H₂O mQ. En paralelo, el vector de expresión pET22b (Novagen) fue digerido ON con las mismas enzimas de restricción y luego tratado por 2 hs a 37°C con fosfatasa Antártica (NEB cat. # M0289). Las enzimas se inactivaron por calentamiento a 80°C durante 5 min y el vector se purificó empleando columnas QIAquick. Las reacciones de ligación se efectuaron ON a 15°C con la enzima T4 DNA Ligase (NEB cat. # M0202S) empleando una relación molar inserto:vector de ~50:1. El producto de la ligación fue dializado contra agua mQ y luego se usó para transformar por electroporación bacterias *E. coli* DH10B. La identidad de las construcciones fue verificada mediante el servicio de secuenciación de ADN de la empresa Macrogen (Korea). El clonado en el vector pET22b añade el hexapéptido His6 en el extremo C-terminal de las proteínas. Los plásmidos conteniendo las GFPs fueron utilizados para transformar la cepa de expresión de *E. coli* BL26.

La expresión de las proteínas se inició a partir de 50 ml de medio LB con el agregado ampicilina 100 µg/ml. Este cultivo fue crecido en agitación a 37°C ON y a la mañana siguiente fue agregado a 1 litro del mismo medio. El cultivo fue crecido a 37°C hasta alcanzar una OD 600 nm de 0.7 y luego se transfirió a un agitador a 25°C. La inducción de las proteínas se realizó mediante el agregado de isopropil tiogalactosido 1 mM (IPTG) y posterior cultivo a 25°C durante 12 hs. Al finalizar la incubación la expresión de ambas proteínas se evidenció por el pronunciado color verde de ambos cultivos. Los mismos fueron centrifugados en frío a 4000 rpm durante 15 min en un rotor GS3 (Sorvall). Los pellets de bacterias se resuspendieron en 100 ml de buffer Tris-HCl 50 mM pH 8, NaCl 150 mM, imidazol 10 mM, lisozima 0.1 mg/ml (SIGMA cat. # L-6876), Triton X-100 0.1 % v/v y PMSF 1 mM. La suspensión fue incubada a RT durante 30 min y luego

enfriada en hielo y sometida a ruptura por ultrasonido (sonicador Omni-Ruptor 4000, 10 pulsos de 20 segundos, 70 % de potencia). La solución fue centrifugada en frío por 20 minutos a 12.000 rpm en un rotor SS34 (Sorvall). Las proteínas fueron purificadas a partir del sobrenadante. El primer paso de purificación se realizó con cromatografía de afinidad por metales inmovilizados (IMAC). Para ello se emplearon columnas empaquetadas en el laboratorio conteniendo 4 ml de matriz cargada con Ni^{2+} (Chelating Sepharose Fast Flow, GE Healthcare Cat. # 17-0575-1). Luego de la siembra, se lavó la columna con 6 volúmenes de Tris-HCl 50 mM pH 8, NaCl 150 mM, imidazol 10 mM, 4 volúmenes de Tris-HCl 50 mM pH 8, NaCl 150 mM, urea 4 M y 4 volúmenes de Tris-HCl 50 mM pH 8, NaCl 150 mM. El lavado con urea limpia considerablemente las muestras, y se puede realizar en este caso porque la urea no afecta el plegado de las *sfGFPs* a temperatura ambiente. Las proteínas fueron eluidas con Tris-HCl 50 mM pH 8, NaCl 150 mM, imidazol 300 mM. La mayor parte eluyó en 2 ml de buffer. A continuación, las proteínas fueron sometidas a un segundo paso de purificación mediante cromatografía de exclusión molecular. Para ello se empleó una columna Superdex 200 (GE Healthcare) con un flujo de 0.7 ml/min de buffer fosfato de sodio 15 mM, pH 7.4, NaCl 150 mM. El perfil de elución fue seguido mediante absorbancia a 280 nm. Este segundo paso de purificación deja a las proteínas en un buffer apropiado para los estudios de dicroísmo circular subsiguientes. La concentración final de proteínas fue determinada con el ensayo de Bradford y su pureza se analizó mediante electroforesis en SDS-PAGE con una concentración de acrilamida-bisacrilamida de 15:1 desarrollado a 110 V durante 3 horas.

XXVIII. Análisis espectroscópicos

La estructura secundaria y la estabilidad térmica de las proteínas fueron analizadas mediante dicroísmo circular (CD) en un espectropolarímetro Jasco J-815 provisto de un termostatizador de efecto Peltier. Se empleó una concentración de proteína de 1 mg/ml y celdas de polarimetría de 1 mm de paso óptico (Hellma). La curva de desplegamiento térmico se efectuó con una rampa de 2°C/min y se siguió la señal de CD a 232 nm. La estabilidad conformacional frente a cloruro de guanidinio (Gdn-HCl) se estudió midiendo la fluorescencia intrínseca del cromóforo de las *sfGFPs* en un espectrofluorímetro Jasco FP-6500. Para ello se empleó una longitud de excitación de 498 nm, y el desplegamiento se siguió midiendo la emisión a 512 nm. El proceso de desplegamiento se siguió también mediante la fluorescencia del único residuo de triptófano que presentan estas proteínas (W57). La longitud de onda empleada para la excitación fue de 295 nm, y se barrió el espectro de emisión de 303 a 400 nm. Los anchos de banda y el voltaje del fotomultiplicador se ajustaron en cada caso. Las muestras de proteínas

fueron incubadas con concentraciones crecientes de GdnHCl durante al menos 96 hs a temperatura ambiente para asegurar que se haya alcanzado el equilibrio. Estos análisis se realizaron a 25°C en buffer fosfato de sodio 50 mM, pH 7.5, NaCl 150 mM, con una concentración de proteína de 0.1 mg/ml. El pH de las soluciones stock de GdnHCl también fue ajustado. Para el estudio de la cinética de plegado y desplegado se empleó un mezclador rápido de dos canales de Applied Photophysics. La jeringa conteniendo la proteína tenía un volumen de 250 μ l, mientras que la que contenía el buffer tenía un volumen 10 veces mayor. El tiempo muerto de este mezclador está cercano a los 2 segundos.

XXIX. Simulaciones de dinámica molecular.

La estructura inicial de la versión superfolder de la GFP (*sfGFP*) fue extraída de la página web de la Protein Data Bank (PDBID: 2B3P; estructura resuelta por cristalografía de rayos-X, 1,4 Å de resolución). Para el armado de la estructura inicial, se conservaron las moléculas de agua presentes en la estructura cristalina. La mutante **C70M** fue generada *in silico* y los átomos de H se agregaron automáticamente usando el programa Chimera para ambos procedimientos. El estado de protonación estándar a pH fisiológico se aplicó a todos los aminoácidos protonables. Las estructuras se solvataron en una caja de solvente explícito octaédrico con moléculas de agua tipo TIP3P, localizando los límites de la caja a 10 Å de la superficie de la proteína. Simulaciones de dinámica molecular (DM) se llevaron a cabo a 1 atm y 300 K (barostato y termostato de Berendsen), usando condiciones periódicas de borde y suma de Ewald para tratar las interacciones de largo alcance con un radio de corte de 10 Å. El algoritmo SHAKE se aplicó a todos aquellos enlaces involucrando átomos de H, permitiendo así el uso de un paso de 2 fs para la integración de las ecuaciones de Newton, parámetros recomendados en el paquete de programas de AMBER. El campo de fuerzas f99SB se utilizó para todos los residuos proteicos. Para equilibrar el sistema se realizaron 8000 pasos de minimización, seguidos de 500 ps de una corrida a volumen constante para termalizar el sistema lentamente hasta 300 K. Finalmente, una corrida de 1 ns de duración se llevó a presión constante para alcanzar una densidad promedio de 1 g mL⁻¹. Las DMs de producción fueron de 40 ns, tiempo suficiente para alcanzar convergencia de todos los parámetros termodinámicos del sistema. Fotos (set de coordenadas y velocidades a un tiempo particular) se colectaron a intervalos de 1 ps y se utilizaron para el análisis posterior.

RESULTADOS

Diseño de la proteína de trabajo.

Elegimos trabajar con una proteína verde fluorescente (GFP) modificada en base a bibliografía y al empleo de herramientas de modelado molecular. Para poder poner a prueba nuestra hipótesis, diseñamos dos ARNm que difieren en cuanto a su composición nucleotídica.

Con el fin de evitar la formación de agregados oligoméricos por la generación de puentes disulfuro entre diferentes GFP, decidimos trabajar con la variante “superfolder” GFP (*sfGFP*) desarrollada por el grupo de trabajo de Waldo⁷⁴. La *sfGFP* es más tolerante a las mutaciones, lo cual sería de suma importancia para introducir secuencias a lo largo de la secuencia aminoacídica. Además, esta variante adquiere su conformación nativa más rápidamente, por lo que se esperaba que la ocupación de sitios de *N*-glicosilación de modo post traduccional resultase disminuida.

Se conoce que la OST puede formar puentes disulfuro transitorios con las proteínas que están siendo translocadas⁷⁵, pudiendo esto aumentar el tiempo en que un sitio pasible de ser *N*-glicosilado ronde el sitio activo de la enzima. Por lo enunciado, el análisis del efecto del uso de codones respecto de la eficiencia de *N*-glicosilación podría complicarse. Fue entonces que decidimos que, si bien la *sfGFP* a pesar de conservar las cisteínas no resulta en agregados, esos residuos debían ser mutados para el fin de nuestro proyecto. En un trabajo publicado⁷⁶, partiendo de una variante de la GFP, se mutan las cisteínas (C48S y C70M) y a la proteína resultante se la denomina cystein free GFP (*cfGFP*), la cual retiene el brillo y no oligomeriza al ser expresada en RE.

Así, a la secuencia nucleotídica codificante para la *sfGFP*, le sumamos las mutaciones correspondientes a las cisteínas y a esa nueva variante la llamamos *cf-sfGFP*. La misma fue construida en dos variantes nucleotídicas: una cuyo ARNm posee codones frecuentes y otra, donde los últimos 63 codones son poco frecuentes. La intención era tener un mensajero con codones que permitieran una lectura ribosomal rápida ó lenta en la distancia estimada a la que el secunón estaría atravesando el sitio activo de la OST.

La primera consideración en la elección de los codones fue la existencia de una rampa de codones poco frecuentes al inicio de los ARNm. Se colocaron 30 codones raros en el comienzo de los mensajeros para evitar que una secuencia que debiera traducirse rápido, lo hiciera a una velocidad menor debido a un colapso entre ribosomas.

A continuación, se completó la secuencia nucleotídica de las variantes *cf-sfGFP* usando como guía la tabla de uso de codones en humano (**ver anexo**). Así, fueron generadas la secuencia

de la *cf-sfGFP* de velocidad “rápida” (con codones frecuentes) y de la *cf-sfGFP* con una ventana de traducción lenta (con codones poco abundantes) al final del ARNm (últimos 63 aminoácidos) (figuras 11 y 12).

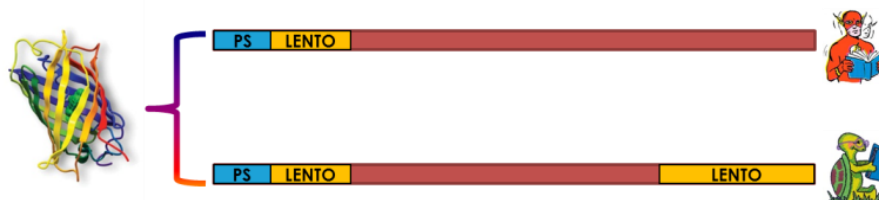


Figura 11. Esquema de las variantes de ARNm empleadas para codificar a la *cf-sfGFP*. PS= péptido señal; LENTO= presencia de codones poco abundantes.

Se sabe que la eficiencia de traducción está modulada por el sesgo en el uso de los codones, el contenido GC y la estructura secundaria del ARNm. Como en su gran mayoría los codones óptimos tienen alto contenido GC, en el caso del ARNm de la proteína “rápida” en vez de optar por el codón óptimo, se eligió al más cercano en abundancia, pero con menor contenido en GC.

Secuencia de ADN de la *cf-sfGFP* versión “fast”:

ATGAGCAAAGGTGAAGAACTATTTACGGGTGTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTC
ATAAATTTTCGGTACGTGGCGAGGGCGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTAC
AACAGGCAAGCTGCCAGTGCCATGGCCAACACTGGTGACAACACTGACATACGGCGTGCGAGATGTTCTCTAGA
TACCCAGACCACATGAAGCAGCAGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAACA
TTTCTTTCAAGGACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACA
GAATTGAGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTTCAA
CTCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACCGGAATTAAGGCTAACTTCAAGATTAGACACAAC
GTGGAGGACGGATCCGTGACGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCCCTGTGCTG
CTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGAGAGACCACATG
GTGCTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTACAAG

Secuencia de ADN de la *cf-sfGFP* versión “slow”:

ATGAGCAAAGGTGAAGAACTATTTACGGGTGTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTC
ATAAATTTTCGGTACGTGGCGAGGGCGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTAC
AACAGGCAAGCTGCCAGTGCCATGGCCAACACTGGTGACAACACTGACATACGGCGTGCGAGATGTTCTCTAGA
TACCCAGACCACATGAAGCAGCAGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAACA
TTTCTTTCAAGGACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACA
GAATTGAGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTTCAA
CTCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACCGGAATTAAGGCTAACTTCAAGATTAGACACAAC
GTGGAGGACGGATCCGTACAACCTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCCGGTACTA
CTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACGTGATCATATG
GTACTACTAGAATTTGTAACGGCGCGGGGTATAACGCATGGTATGGATGAACTATATAAA

Figura 12. Secuencias nucleotídicas de ambas versiones de la proteína de trabajo. Se muestra subrayado el sitio de restricción BamHI y, en negrita se indican los segmentos moduladores de la velocidad.

Una vez obtenidas las secuencias nucleotídicas que codificarían para nuestras proteínas, se empleó el programa Vienna RNA para visualizar las estructuras secundarias predichas para los ARNm. La intención era tener dos ARNm con diferente uso de codones que no difirieran demasiado en la formación de estructura secundaria. Además, se usó un algoritmo de GenScript (OptimumGene) para analizar el contenido de GC, sabiendo que el ideal está comprendido entre 30-70% (**Figura 13**).

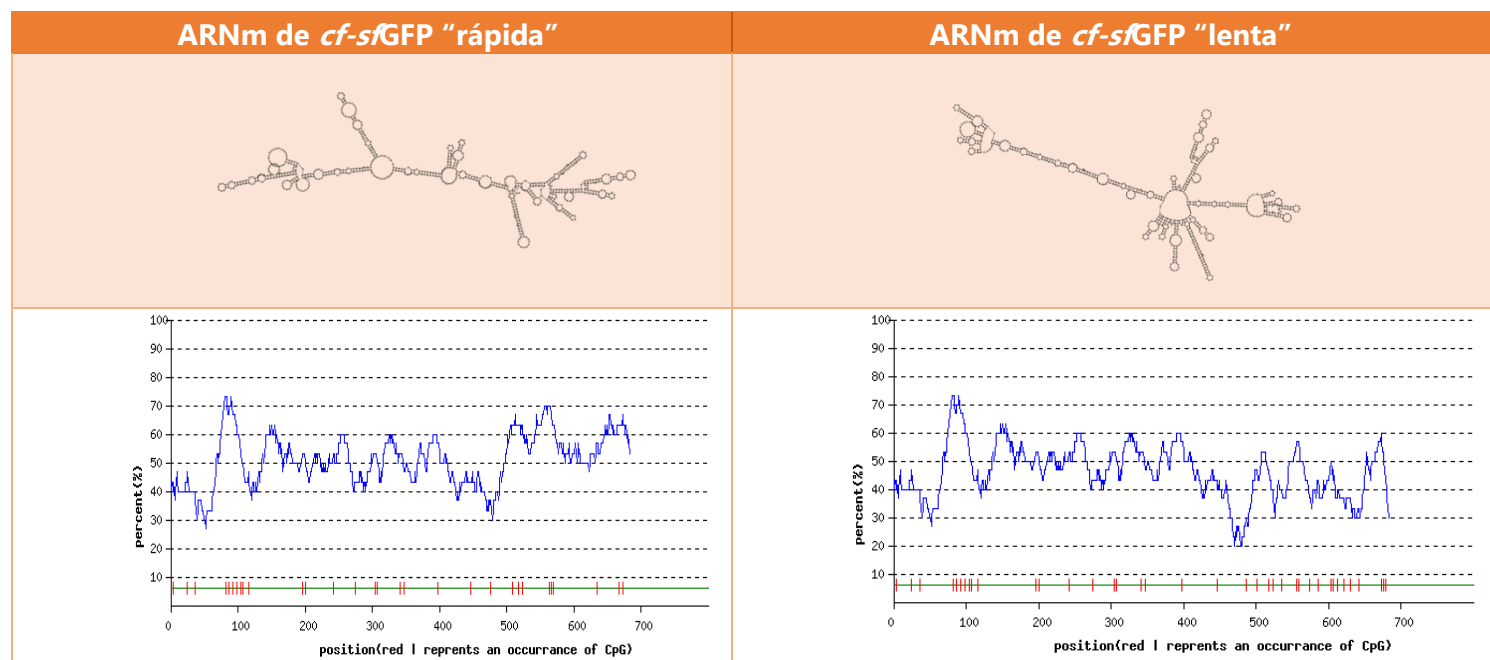


Figura 13. En la zona superior se observa la estructura predicha para los ARNm según el programa Vienna RNA, mientras que en la porción inferior se muestra el porcentaje de contenido GC a lo largo de los ARNm.

Una vez diseñado el ARNm en sus dos variantes, se encargó la síntesis del ADN correspondiente a GenScript. Pero, en lugar de mandar a sintetizar el ADN que daría como resultado de su expresión a la proteína lenta, se introdujo un sitio de restricción BamHI justo antes de los últimos 63 aminoácidos en la secuencia que daría la versión "rápida", de modo que pidiendo la secuencia lenta con ese mismo sitio de restricción luego se pudiera clonar el fragmento "lento" y obtener así ambas variantes.

Clonado del ADN en vectores de expresión en células de mamífero.

El ADN de la secuencia rápida fue amplificado por PCR con cebadores diseñados con sitios de restricción en sus extremos. El producto de la reacción fue luego digerido y finalmente ligado con vectores de expresión en mamíferos (pcDNA3.1), que poseen señales que permiten el direccionamiento a retículo endoplásmico, citoplasma ó medio extracelular (**figura 14**).

Obtenidos los vectores de la proteína “fast”, se procedió a generar las variantes “slow” a partir de la digestión de los mismos con BamHI y EcoRI, seguido de la ligación de la secuencia lenta levantada por PCR con cebadores en cuyos extremos estaban los mismos sitios de corte.

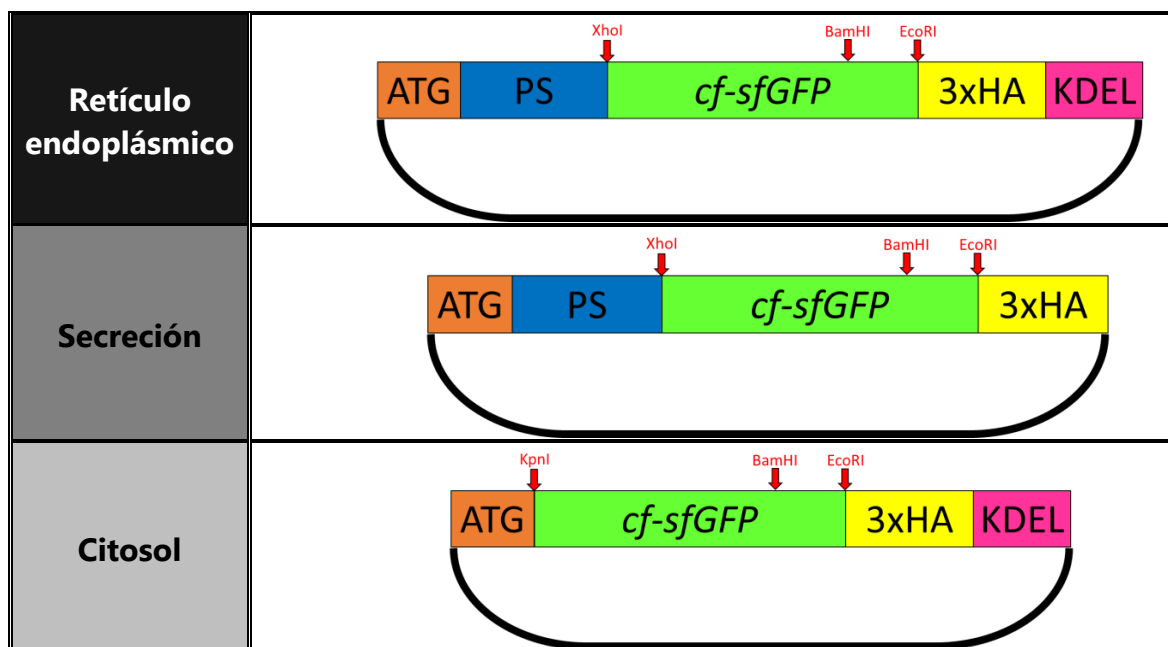


Figura 14. Esquema de la *cf-sfGFP* clonada en vectores de direccionamiento de expresión en mamífero.

Una vez obtenidas las variantes rápidas/lenta en los diferentes vectores de direccionamiento, se transfectaron células HEK293 para evaluar la correcta localización de la proteína en el compartimiento subcelular al cual se la hubiere dirigido, su fluorescencia (**figura 15**) y la expresión de la misma por western blot (**figura 16**). Era importante verificar que la proteína se estuviera expresando y que fuese fluorescente a pesar de las modificaciones realizadas en los codones. La proteína lenta, no sólo cuenta con 30 codones poco abundantes al inicio de su secuencia, sino que además posee los 63 codones del segmento lento al final, lo cual podría haber modificado su nivel de expresión. Sin embargo, se obtuvieron cantidades comparables de ambas versiones.

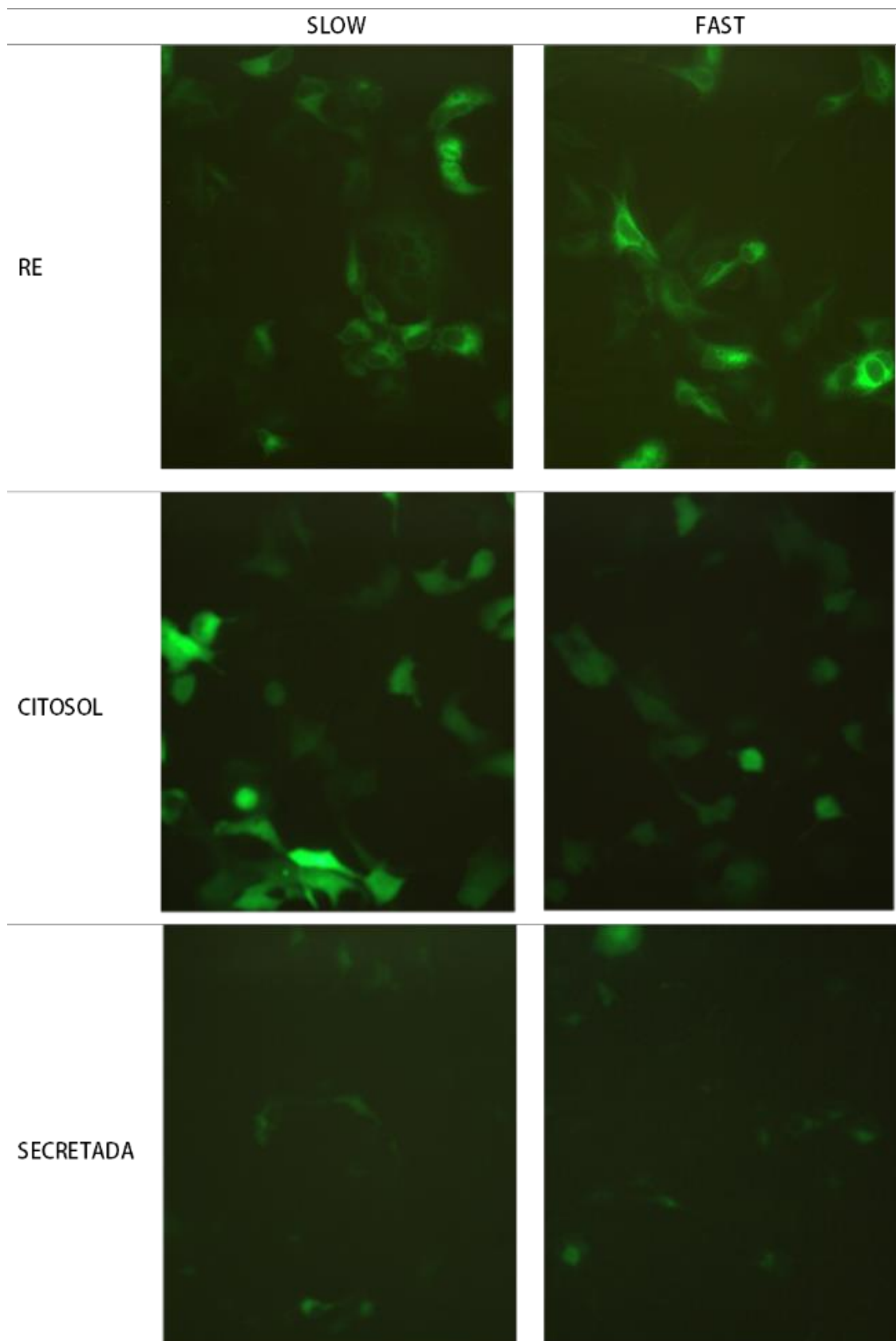


Figura 15. Imágenes tomadas con el microscopio de fluorescencia (A=400x) de las variantes de la *cf-sfGFP* clonadas en los distintos vectores de direccionamiento de expresión en mamífero, transfectadas en células HEK293.

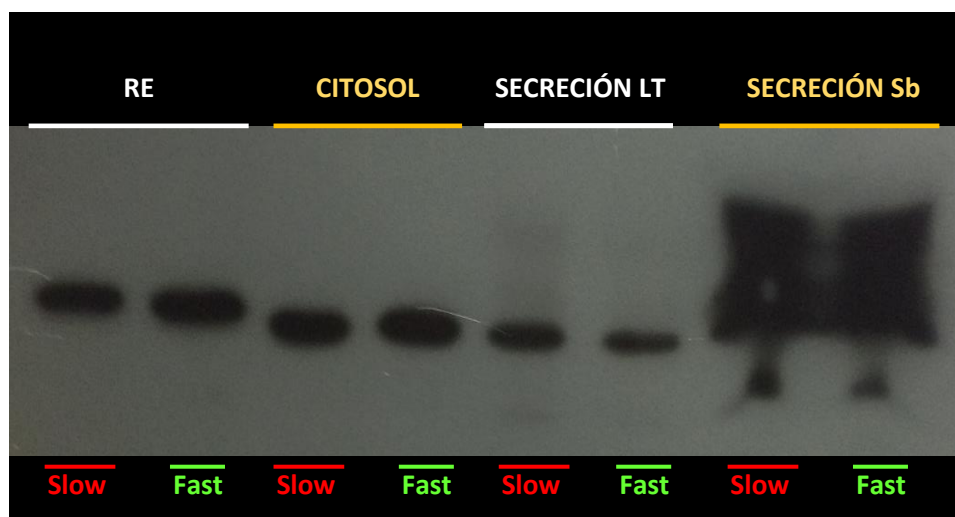


Figura 16. Resultados del western blot realizado con muestras proteicas provenientes de la transfección en células HEK293 de las variantes de la *cf-sfGFP* clonadas en los distintos vectores de direccionamiento de expresión en mamífero. LT: lisado total; Sb: sobrenadante de cultivo.

Mutantes de *N*-glicosilación

i) Posición de los secuencias.

A partir de las construcciones lenta y rápida, se generaron por mutagénesis dirigida sitios de *N*-glicosilación en diferentes posiciones de la secuencia proteica. Se trabajó sobre el vector de direccionamiento a RE, ya que es en dicho compartimento subcelular donde puede ocurrir la *N*-glicosilación.

Primero se analizó la secuencia aminoacídica en busca de residuos Asn, Ser ó Thr existentes, para que sólo hubiera que realizar una mutación puntual para obtener al secunón (Asn-X≠Pro-Ser/Thr). Una vez anotados todos los secunones posibles, se usó el programa Chimera para visualizar la localización de la asparagina a ser modificada y la factibilidad de colocar allí una modificación tan voluminosa e hidrofílica como es un *N*-glicano. Nuestro grupo de trabajo publicó recientemente un trabajo que muestra la existencia de una enorme presión de selección negativa sobre aquellos secunones que quedan en posiciones ocultas de las proteínas⁷⁷. Por este motivo, se descartaron los residuos Asn ocultos y aquellos cambios de los que se infirió que podrían romper con la estructura de la proteína no glicosilada. Dentro de ese conjunto de posibles candidatos, se seleccionaron secunones en diferentes posiciones a lo largo de la secuencia aminoacídica (**figura 17**), se diseñaron los cebadores correspondientes para realizar las mutaciones y se ordenó su síntesis.

MDSKGSSQKGSRLLLLVSNLLCQGVVSTPLESKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTCLKF
 ISTTGKLPVPWPTLVTTLTYGVMFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRI
 ELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHY
 LSTQSVLSKDPNEKRDHMLLEFVTAAGITHGMDELYKEFYYPDVPDYAGYPYDVPDYAGSYDVPDYAAKDEL

Figura 17. Secuencia aminoacídica de la proteína de trabajo **ATG-SP prolactin -cf-sfGFP-3xHA-KDEL**. Se muestran en diferentes colores los sitios elegidos para generar los secuones por mutagénesis. La “pausa” (aminoácidos codificados por codones poco frecuentes) se encuentra subrayada y resaltada en amarillo.

En la **figura 18** se muestra la estructura de la GFP y la ubicación de los secuones.

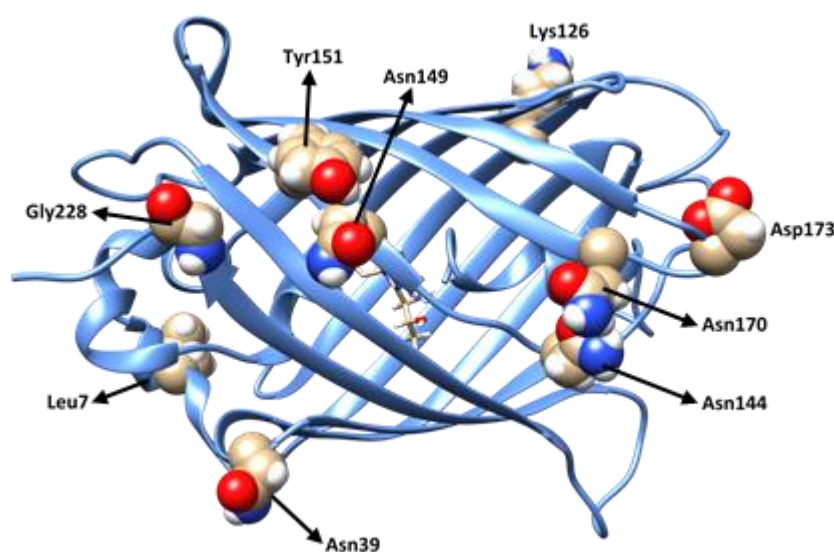


Figura 18. Se indica la posición de la asparagina de cada secuón a generar dentro de la estructura de la GFP.

ii) Generación de sitios de N-glicosilación y transfecciones.

Se generaron por mutagénesis dirigida por PCR, sitios de N-glicosilación (secuones) con serina en la tercera posición, ubicados a diferentes distancias respecto del segmento modulador de la velocidad (los últimos 63 codones).

Las mutantes de N-glicosilación en sus versiones rápida/lenta (con codones abundantes ó poco frecuentes a la distancia estimada en la que el secuón estaría atravesando el sitio activo de la OST) fueron obtenidas y transfectadas en líneas celulares humanas. Se verificó bajo microscopio la fluorescencia de todas las mutantes luego de 24 hs de transfección (**figura 19**) y se procedió a la lisis celular para obtener un extracto proteico, que sería luego hervido y guardado en alícuotas a -80°C.

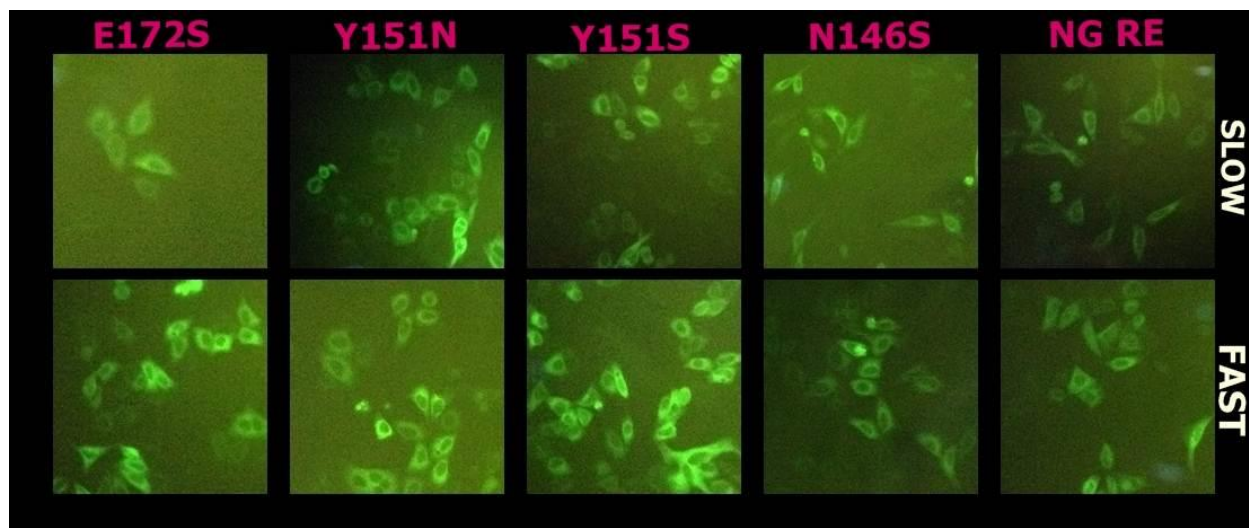


Figura 19. Imágenes tomadas con el microscopio de fluorescencia de las variantes de la *cf-sfGFP* clonadas en el vector de expresión en mamífero que dirige al RE, transfectadas en células HeLa.

Luego, mediante western blot y posterior análisis con el programa Image J, se cuantificó la relación proteína glicosilada sobre proteína total ($G/(G+NG)$). Se consideró relevante comparar los cocientes en aquellos casos donde la cantidad de proteína total fuese similar. Así, se observó un aumento del 11% en la *N*-glicosilación de la variante lenta de la mutante N146S respecto de su versión rápida. Por el contrario, en el caso de la mutante E172S observamos una disminución de la *N*-glicosilación de la variante lenta respecto de la rápida en un 6% (figura 20 y tabla 3).



Figura 20. Western blot de las mutantes N146S y E172S de la *cf-sfGFP*, clonadas en el vector de expresión en mamífero que dirige al RE y transfectadas en células HeLa.

	G/(G+NG)	
	slow	fast
N146S	0,51	0,46
E172S	0,65	0,69

Tabla 3. Cuantificación del western blot de la figura 20.

Los cálculos para las mutantes Y151N, Y151S y G228N+T230S no se muestran porque no presentaban un patrón de doble banda (**figura 21**). En estos casos la *N*-glicosilación tuvo lugar con una eficiencia del 100 % o muy cercana.

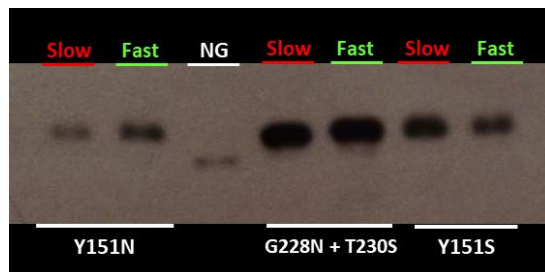


Figura 21. Western blot de las mutantes indicadas de la *cf-sfGFP*, clonadas en el vector de expresión en mamífero que dirige al RE transfectadas en células HeLa.

Para determinar si en estas mutantes existía algún momento en el que hubiera proteína no glicosilada, efectuamos una curva de tiempo (**figura 22**) en la que variamos el momento post transfección en el cual procedimos a lisar las células (t=0, 4, 8 y 16 hs).

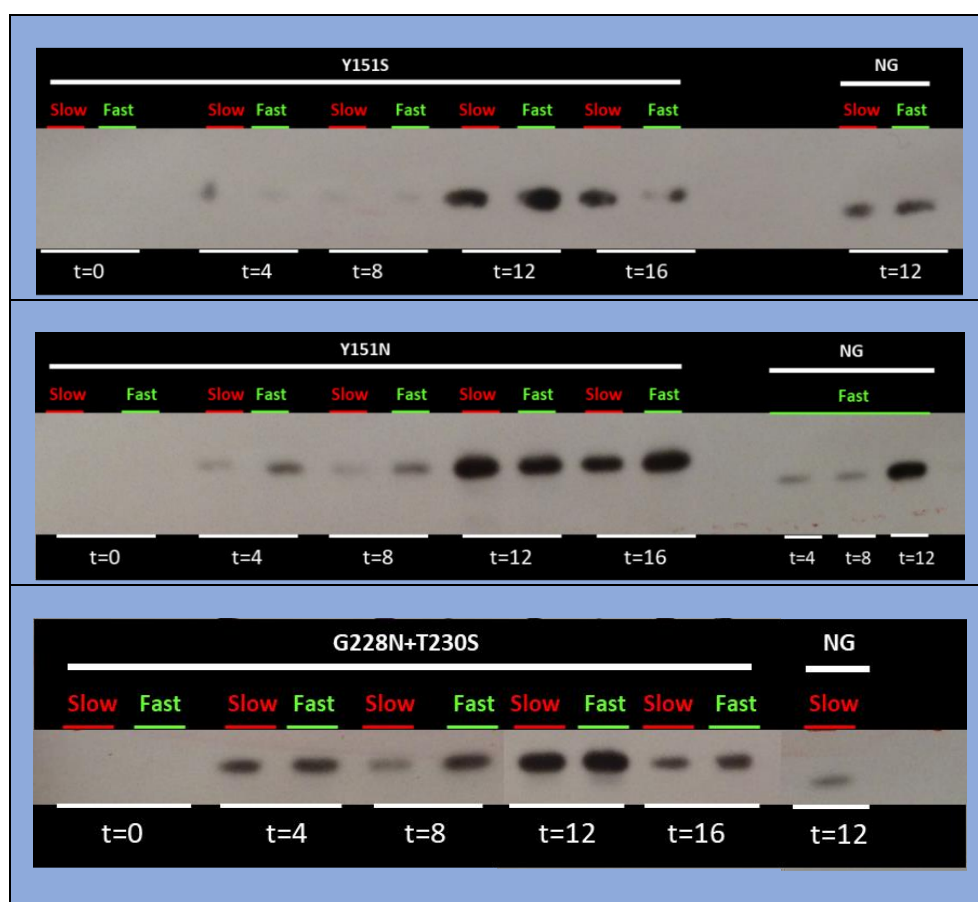


Figura 22. Western blot de las mutantes Y151S, Y151N, G228N+T230S de la *cf-sfGFP* y de la variante no glicosilada (NG), clonadas en el vector de expresión en mamífero que dirige al RE transfectadas en células HeLa. El tiempo "t" se indica en hs post transfección.

Luego de realizar el western blot, como puede observarse en la **figura 22**, las mutantes G228N+ T230S, Y151N e Y151S resultaron totalmente glicosiladas independientemente del tiempo analizado.

Se realizó la misma curva de tiempo para las mutantes que presentaban una doble banda y, como puede observarse en la **figura 23**, el patrón se detecta desde las 4 hs y se conserva hasta las 16 hs. Estos resultados indican que el tiempo característico del proceso de *N*-glicosilación es mucho más corto que los tiempos empleados en estos experimentos.

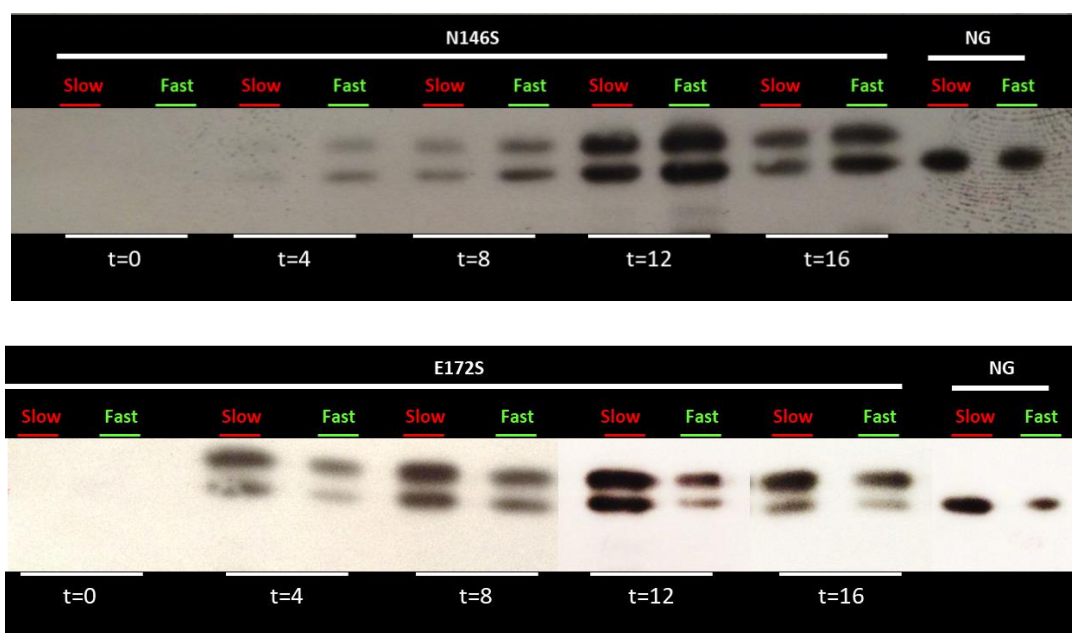


Figura 23. Western blot de las mutantes N146S y E172S de la *cf-sfGFP* y de la variante no glicosilada (NG), clonadas en el vector de expresión en mamífero que dirige al RE transfectadas en células HeLa. El tiempo “t” se indica en hs post transfección.

A continuación, se realizó una nueva transfección en células HEK293 (**figura 24**). Con los lisados celulares obtenidos luego de 13 hs, se realizaron los western blots correspondientes y las bandas fueron cuantificadas. Se consideró relevante comparar los cocientes en aquellos casos donde la cantidad de proteína total fuese similar.

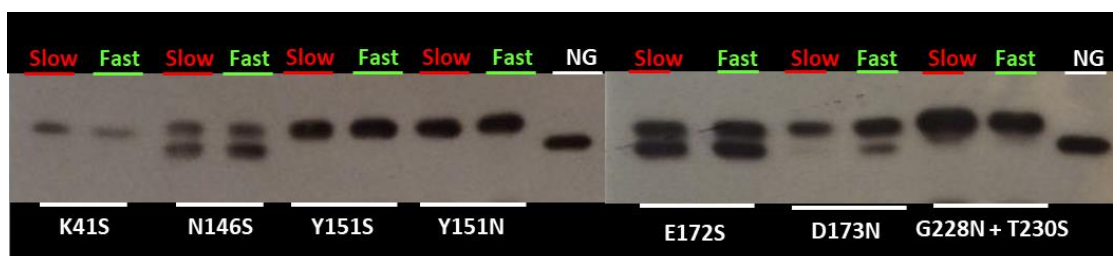


Figura 24. Western blot de las mutantes indicadas de la *cf-sfGFP*, clonadas en el vector de expresión en mamífero que dirige al RE transfectadas 13hs en células HEK293.

Puede observarse que, al igual que en la línea celular HeLa, las mutantes Y151S, Y151N, G228N+T230S muestran una única banda correspondiente a proteína completamente *N*-glicosilada, mientras que N146S y E172S muestran un patrón de doble banda. En este caso hay dos mutantes nuevas: por un lado, la K41S se muestra totalmente *N*-glicosilada mientras que la D173N se presenta tanto en su versión *N*-glicosilada como en la forma no glicosilada.

Como puede observarse en la **tabla 4**, esta vez al cuantificar con el Image J, la mutante E172S no muestra diferencias entre las versiones slow y fast, mientras que la mutante N146S sigue mostrando un aumento (18% vs 11% en HeLa) en la *N*-glicosilación en la versión lenta respecto de la rápida.

	G/(G+NG)	
	slow	fast
N146S	0,45	0,38
E172S	0,52	0,51

Tabla 4. Cuantificación del western blot de la figura 24.

En otro western blot realizado con lisados celulares provenientes de la transfección por 24hs de células HEK293 con la mutante E172S (**figura 25**), los resultados fueron diferentes, mostrando una mejora del 13% en la eficiencia de *N*-glicosilación en el caso de la versión rápida respecto de la lenta (**tabla 5**).

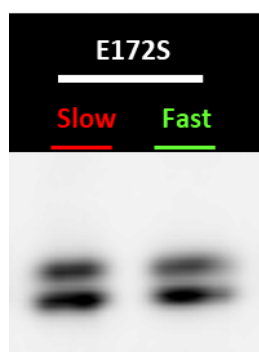


Figura 25. Western blot de las mutantes indicadas de la *cf-sfGFP*, clonadas en el vector de expresión en mamífero que dirige al RE transfectadas en células HEK293.

	G/(G+NG)	
	slow	fast
E172S	0,45	0,51

Tabla 5. Cuantificación del western blot de la figura 25.

En el caso de la mutante D173N se realizó otro western blot (**figura 26**) para obtener una intensidad total de proteína (G+NG) comparable entre las versiones lenta/rápida y no hubieron diferencias apreciables al cuantificar (**tabla 6**).



Figura 26. Western blot de la mutante D173N de la *cf-sfGFP*, clonadas en el vector de expresión en mamífero que dirige al RE, transfectada en células HEK293.

	G/(G+NG)	
	slow	fast
D173N	0,78	0,76

Tabla 6. Cuantificación del western blot de la figura 26.

COS-7

Se transfectaron las mutantes D173N y N146S en células COS-7, derivadas de riñón de mono, dónde el uso de codones es bastante similar al uso de codones en humanos (**figura 27**).



Figura 27. Western blot de las mutante D173N y N146S de la *cf-sfGFP*, clonadas en el vector de expresión en mamífero que dirige al RE, transfectadas en células COS-7.

Puede apreciarse un incremento del 16 % en la glicosilación de la variante lenta de la mutante N146S respecto de su versión rápida en las células COS-7, mientras que la mutante D173N muestra una diferencia cercana al 4 % entre las variantes, resultados esperables por lo antedicho respecto del uso de codones.

	G/(G+NG)	
	slow	fast
D173N	0,80	0,83
N146S	0,43	0,37

Tabla 7. Cuantificación del western blot de la figura 27.

Dada la ausencia de diferencia entre slow y fast ó la presencia de cambios muy pequeños en el grado de *N*-glicosilación, pensamos que eso podía deberse a que la identidad de los secuencias con los que estábamos trabajando los hacía ser muy eficientes en cuanto a su *N*-glicosilación, perdiéndonos así la visualización del efecto de la modulación de la velocidad de traducción sobre el proceso. Entonces, se diseñaron cebadores para mutar al residuo “Y” (el siguiente a la serina) con el fin de empeorar la eficiencia del secuéon. Se ha descrito que colocar ácido glutámico (E) en la posición “Y” disminuye la eficiencia de *N*-glicosilación⁷⁸. Teniendo en cuenta eso, se obtuvieron las siguientes mutantes:

- **NS**= N146S+S147E
- **NISE**= Y151N+T153S+A154E
- **DV**= D173N+V175E
- **KID**= K126N+I128S+D129E. Esta es una mutante totalmente nueva.

Una vez generadas las 8 nuevas mutantes fueron transfectadas en células HEK293. A las 13hs se obtuvieron los lisados celulares y se realizó un western blot (**figura 28**), que fue cuantificado con el programa Image J (**tabla 8**). Los resultados para la mutante NISE se mostrarán más adelante.

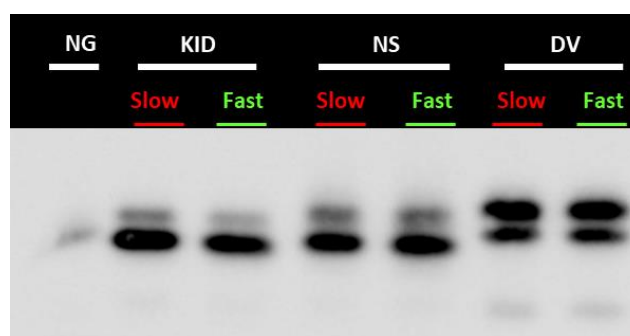


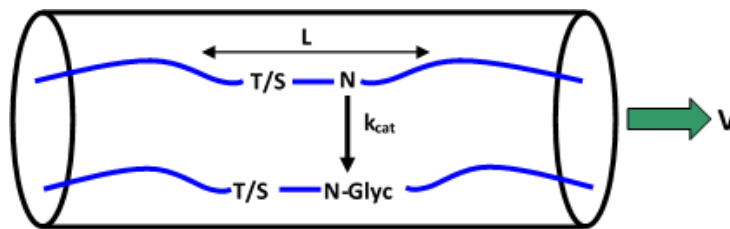
Figura 28. Western blot de las mutantes KID, NS, DV de la *cf-sfGFP* y de la variante no glicosilada (NG), transfectadas en células HEK293 por 13hs.

	G/(G+NG)	
	slow	fast
KID	0,21	0,19
NS	0,31	0,30
DV	0,63	0,60

Tabla 8. Cuantificación del western blot de la figura 28.

Posiblemente la no variación de los secuencias KID, NS y DV podría deberse a que la identidad de los mismos fue empeorada en gran magnitud y que, por consiguiente, por más que los sitios estuviesen en un contexto lento al alcanzar el sitio activo de la OST, no se pudiera observar la mejora en la *N*-glicosilación. Para comprender de forma semicuantitativa la relación entre velocidad de traducción y eficiencia de *N*-glicosilación se desarrolló un modelo matemático muy simple que describe el proceso. Este modelo postula que la ocupación co-traduccional ocurrirá principalmente mientras el secunón se encuentre en la proximidad del sitio activo de la OST. Asumiendo que la catálisis puede ocurrir dentro de un determinado rango de distancias *L* y que la proteína ingresa al RE a medida que es traducida a una velocidad *V*, esto indica que el tiempo de residencia disponible para la catálisis será:

$$t_{cat} = L / V$$



La suma de las fracciones de proteína glicosilada (G) y no glicosilada (NG) está relacionada por:

$$1 = [G] + [NG]$$

La velocidad de transferencia del glicano es una reacción de segundo orden, que depende de la fracción de sustrato no glicosilado ([NG]), de la constante catalítica *k_{cat}* y de la concentración de lípido-glicano dador. Para simplificar, se supone el lípido-azúcar dador se encuentra saturando la enzima. Esta velocidad está dada entonces por:

$$d[NG] / dt = - k_{cat} \cdot [NG]$$

Integrando esta ecuación para el período de permanencia en el sitio catalítico *t_{cat}*, tenemos que la proporción de proteína no glicosilada es:

$$[NG] = \exp (-k_{cat} \cdot t_{cat}) = \exp (-k_{cat} \cdot L / V)$$

Y la fracción de proteína glicosilada es:

$$[NG] = 1 - \exp(-k_{cat} \cdot L / V)$$

Esta ecuación indica que a mayor eficiencia catalítica (secuencias muy eficientes), o a una menor velocidad de traducción, se verá un incremento en la eficiencia de ocupación del sitio. Para comprender mejor el proceso se realizó una simulación de la dependencia de la ocupación de los sitios en función de la velocidad de pasaje por el sitio activo para tres tipos de secuencias posibles (**figura 29**).

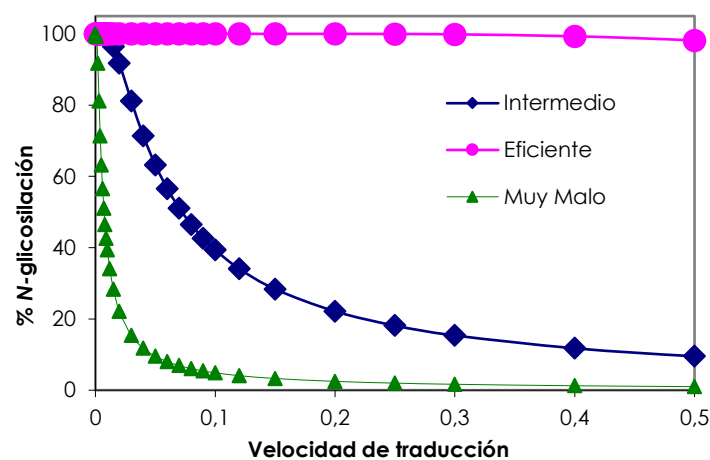


Figura 29. El gráfico surge de la simulación de un modelo matemático que relaciona la eficiencia de N-glicosilación con la velocidad de traducción.

La relación entre estas variables claramente es no lineal. Se puede observar en la simulación que para secuencias que son muy buenas la ocupación es cercana al 100 %, independientemente de la velocidad de traducción. Algo similar ocurre para secuencias que son muy malos, los cuales sólo se ocupan apreciablemente si la velocidad de traducción es muy baja. Existiría un rango de secuencias intermedios, cuya ocupación podría verse afectada apreciablemente al cambiar la velocidad de traducción.

Variantes KID. En el caso del sitio KID se observó un cambio muy pequeño (**figura 28 y tabla 8**) sobre un basal de glicosilación muy bajo, por lo que pensamos que, si mejorábamos su identidad, quizás podríamos llegar a un rango dinámico tal que el cambio entre rápido y lento fuera mucho más apreciable. Para eso, podíamos dejar la D129 original de la *cf-sfGFP* ó

cambiarla por alanina (D129A). Entonces, se generaron las nuevas construcciones por PCR inversa, K1 y K2 (figura 30).

K1: K126N-I128S (AAG126AAC)I128S(ATT128AGC)

MDSKGSSQKGSRLLLLLLVVSNLLLCQGVVSTPLESKGEELFTGVVPILVELDGDVNGHKFSVRGEGEG
DATNGKLTCLKFISTTGKLPVPWPTLVTTLTYGVMFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIEL**NGSD**FKEDGNILGHKLEYNFNShNVYITADKQKNGIKANFKIRHNVE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMLLEFVTAAGITHGMDELYKEF
YPYDVPDYAGYPYDVPDYAGSYDPYDVPDYAAKDEL

K2: K126N-I128S-D129A (AAG126AAC)I128S(ATT128AGC)D129A (GAC129GCC)

MDSKGSSQKGSRLLLLLLVVSNLLLCQGVVSTPLESKGEELFTGVVPILVELDGDVNGHKFSVRGEGEG
DATNGKLTCLKFISTTGKLPVPWPTLVTTLTYGVMFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIEL**NGSA**FKEDGNILGHKLEYNFNShNVYITADKQKNGIKANFKIRHNVE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMLLEFVTAAGITHGMDELYKEF
YPYDVPDYAGYPYDVPDYAGSYDPYDVPDYAAKDEL

Figura 30. Secuencia aminoacídica de las nuevas mutantes generadas, K1 y K2.

Las nuevas mutantes generadas fueron transfectadas en células HEK293. Con los lisados celulares se realizó un western blot (figura 31), donde se aprecia que se logró una mejora en la eficiencia de *N*-glicosilación respecto de la versión KID. Se calculó un 16% de aumento en la *N*-glicosilación de la variante lenta respecto de la rápida en la mutante K2, mientras que en la mutante K1 la diferencia alcanza el 40% (tabla 9).

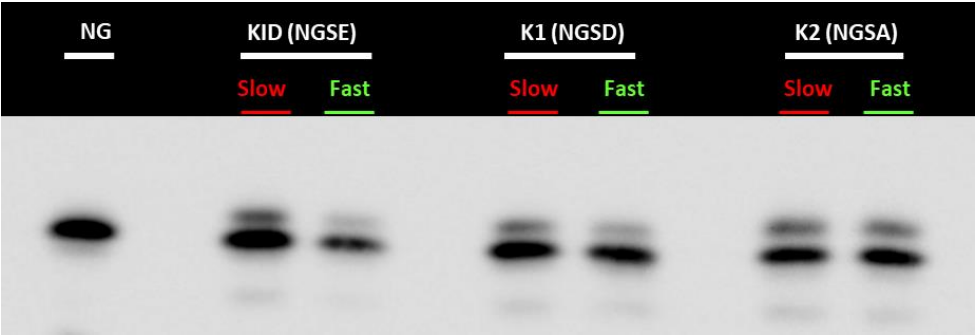


Figura 31. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

	G/(G+NG)	
	slow	fast
K1	0,28	0,20
K2	0,36	0,31

Tabla 9. Cuantificación del western blot de la figura 31.

HEK293 KNOCKOUT

La edición del genoma en células humanas proporciona una poderosa herramienta para crear líneas celulares knock-out (KO). En un trabajo⁷⁹ publicado por el grupo del Dr. Reid Gilmore, mediante la aplicación de la técnica CRISPR/Cas9 se generaron células HEK293 que no expresan la subunidad STT3A (**A^{-/-}**) ó la STT3B (**B^{-/-}**) de la oligosacariltransferasa. Nos resultó interesante acceder a dichas líneas KO dado que en nuestro trabajo queríamos estudiar la *N*-glicosilación que ocurre de modo co-traduccional (proceso llevado a cabo principalmente por la subunidad STT3A). Se esperaba que mediante la transfección de la línea HEK293 STT3B^{-/-}, la *N*-glicosilación observada fuera principalmente aquella que se realizó de modo co-traduccional. Entonces, nos contactamos con el Dr. Gilmore, quien accedió muy gentilmente a enviarnos las líneas HEK293 KO generadas por su grupo.

Una vez que arribaron las células, luego de un mes de recuperación de las mismas, se pudo generar un stock de trabajo para guardar en el tanque de nitrógeno.

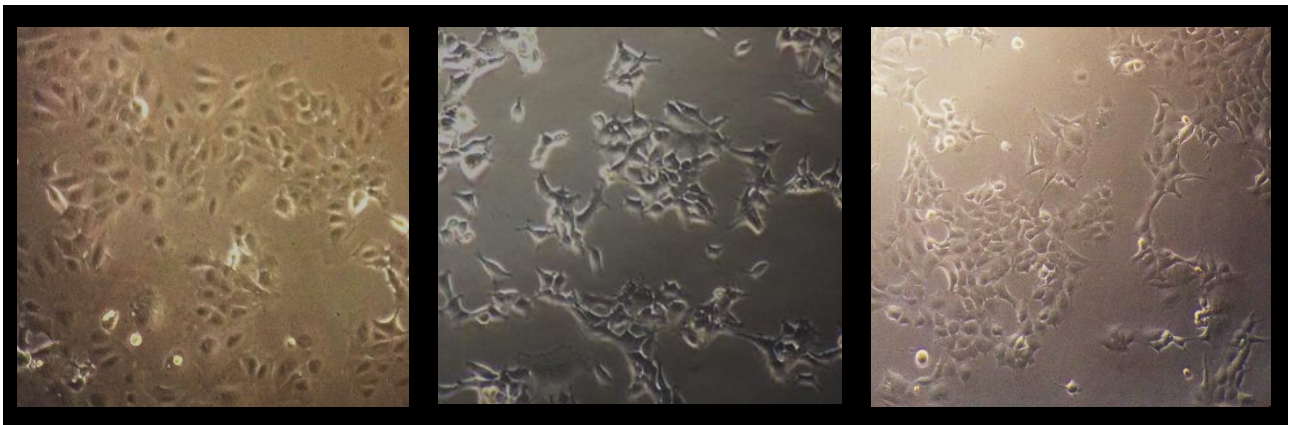


Figura 32. Imágenes tomadas de las células HEK293 enviadas por el Dr. Gilmore. De izquierda a derecha: wild type, A^{-/-}, B^{-/-}.

La mutante N146S fue transfectada en las líneas knockout provistas por el Dr. Gilmore y los resultados pueden observarse en la **figura 33**.

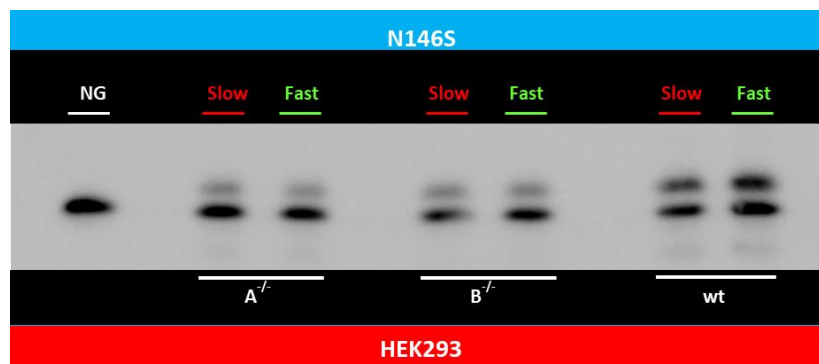


Figura 33. Western blot realizado con lisados celulares de células HEK293 transfectadas con la mutante N146S ó con el control no glicosilado (NG).

Una observación general en esta serie de experimentos es que el grado de ocupación en las células wt siempre es mayor o igual al observado en cualquiera de las líneas KO. Esto es algo lógico dado que la línea wt dispone de ambas maquinarias enzimáticas. Al cuantificar la **figura 33** con el Image J, se obtuvo un 18 % de aumento en la glicosilación de la variante lenta respecto de la rápida en el caso de las células KO para la subunidad STT3B (aquellas que sólo podrían modificar co-traduccionamente a los secunones), mientras que las variantes transfectadas en células HEK293 KO para las STT3A no muestran diferencias (**tabla 10**). Debe destacarse que ese mismo porcentaje de aumento es el que se había medido anteriormente en el caso de las HEK293 de nuestro laboratorio. Este resultado estaría sugiriendo que la subunidad A de la oligosacariltransferasa podría estar supliendo el rol de la B.

N146S	G/(G+NG)	
	slow	fast
A ^{-/-}	0,25	0,25
B ^{-/-}	0,33	0,28
wt	0,45	x

Tabla 10. Cuantificación del western blot de la figura 33.

Los resultados de la transfección en células silvestre no fueron analizados porque la cuantificación total de proteína resultó no ser comparable. Por este motivo, se realizó una nueva transfección de 48 hs (**figura 34**). En este caso, pueden compararse wt_{fast} vs A^{-/-}_{fast}, todas las variantes slow entre sí, A^{-/-}_{fast} vs A^{-/-}_{slow} y wt_{fast} vs wt_{slow}. Lo que se observa es que en cuanto a nivel de glicosilación el orden es wt>B^{-/-}>A^{-/-} (**tabla 11**).

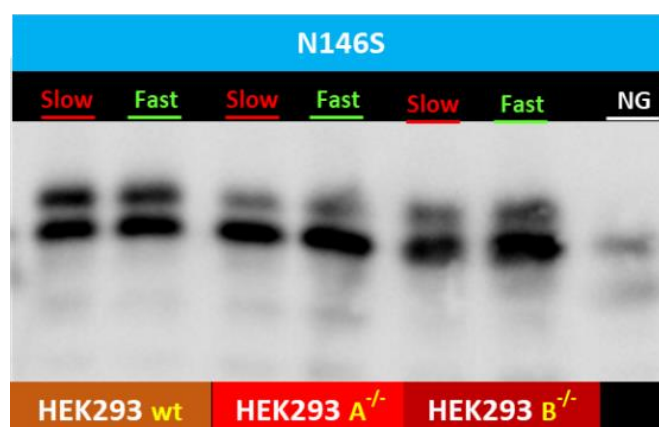


Figura 34. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

N146S	G/(G+NG)
	fast
wt	0,44
A ^{-/-}	0,30
B ^{-/-}	x

N146S	G/(G+NG)
	slow
wt	0,46
A ^{-/-}	0,28
B ^{-/-}	0,35

N146S	G/(G+NG)	
	slow	fast
wt	0,46	0,44
A ^{-/-}	0,30	0,28
B ^{-/-}	0,35	x

Tabla 11. Cuantificaciones del western blot de la figura 34.

Al transfectar la mutante NISE en las distintas líneas HEK293 y realizar el western blot correspondiente, se pudo observar que en las células KO para la STT3A aparece una banda no glicosilada, ausente en las otras dos líneas (**figura 35**). Esto nos estaría indicando que al faltar la subunidad que realizaría la modificación de modo co-traduccional, la STT3B no puede suplir completamente su falta y comienza a haber una banda de proteína sin glicano.

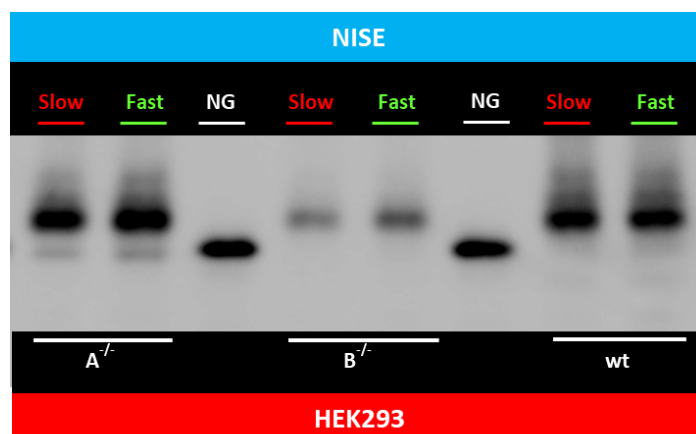


Figura 35. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

NISE	G/(G+NG)	
	slow	fast
wt	1,00	1,00
A ^{-/-}	0,89	0,86
B ^{-/-}	1,00	1,00

Tabla 12. Cuantificaciones del western blot de la figura 35.

Se repitieron los western blots y los resultados fueron reproducidos, como se observa a en la **figura 36 y tabla 13**.

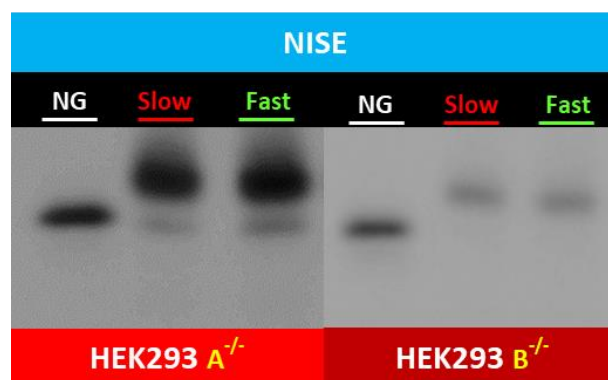


Figura 36. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

NISE	G/(G+NG)	
	slow	fast
wt	1,00	1,00
A ^{-/-}	0,88	0,86
B ^{-/-}	1,00	1,00

Tabla 13. Cuantificaciones del western blot de la figura 36.

A continuación, las mutantes E172S y DV fueron transfectadas en las líneas HEK293 wt, STT3A^{-/-} y STT3B^{-/-}, durante 48 hs. Luego, con los extractos celulares se realizaron los western blots (figura 37) y análisis correspondientes.



Figura 37. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

En el caso de la mutante DV, de la cuantificación con el ImageJ resultó no ser comparable el total de proteína entre slow y fast para cada línea celular, pero sí pudo compararse a las versiones fast entre líneas y a las lentas entre wt y B^{-/-}. Como se aprecia en la **tabla 14**, tanto en las células B^{-/-} como en las A^{-/-}, la glicosilación es menor que en las células wt, y comparando A^{-/-} vs B^{-/-}, las A^{-/-} glicosilan más. Esto último también se evidencia para la mutante E172S (**tabla 15**).

DV	G/(G+NG)	DV	G/(G+NG)
	fast		slow
wt	0,50	wt	0,51
A ^{-/-}	0,39	A ^{-/-}	x
B ^{-/-}	0,31	B ^{-/-}	0,33

Tabla 14. Cuantificación del western blot de la figura 37.

E172S	G/(G+NG)	E172S	G/(G+NG)
	slow		fast
wt	x	wt	x
A ^{-/-}	0,32	A ^{-/-}	0,37
B ^{-/-}	0,27	B ^{-/-}	0,24

Tabla 15. Cuantificación del western blot de la figura 37.

En otro western blot, se pudieron comparar los resultados de las variantes de velocidad para la E172S (**figura 38**), tanto en células A^{-/-} como en las wt. Se puede apreciar que, en el caso de las células silvestres, la variante rápida muestra un aumento en la eficiencia de N-glicosilación respecto de la lenta, mientras que en las células que carecen de la subunidad STT3A no existen diferencias (**tabla 16**). Esto implicaría que el cambio que se observa en las células wt se debe a un proceso co-traducciona

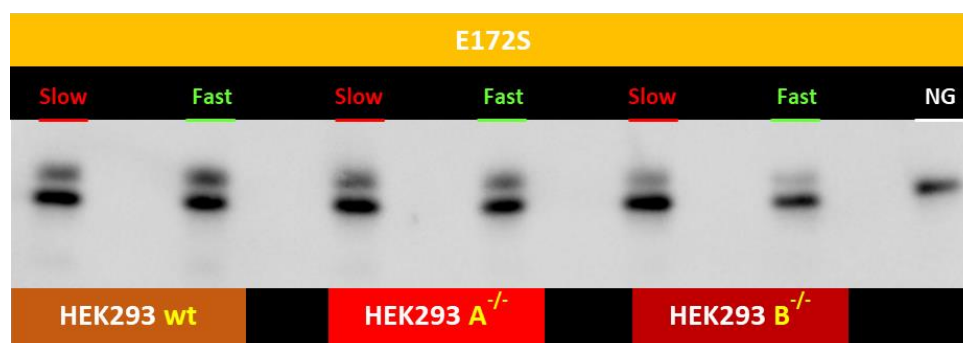


Figura 38. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

E172S	G/(G+NG)		E172S	G/(G+NG)	E172S	G/(G+NG)
	slow	fast		slow		fast
wt	0,37	0,44	wt	0,37	wt	0,44
A ^{-/-}	0,37	0,37	A ^{-/-}	0,37	A ^{-/-}	0,37
B ^{-/-}	x	x	B ^{-/-}	0,27	B ^{-/-}	x

Tabla 16. Cuantificación del western blot de la figura 38.

En el caso de la mutante N-terminal, G228N+T230S, lo que se observa cualitativamente en la **figura 39** es que en las células que carecen de la subunidad STT3B es donde hay un mayor porcentaje de proteína no glicosilada. Este secuón se encuentra a menos de 55 residuos del extremo c-terminal de la proteína y, es de esperar que su ocupación se vea más afectada cuando falta la subunidad STT3B. Sin embargo, es de destacar que la STT3A por sí sola es capaz de ocupar parcialmente dicho secuón.

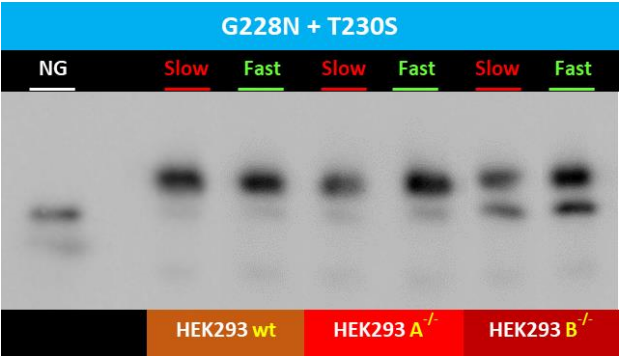


Figura 39. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

G228N + T230S	G/(G+NG)	
	slow	fast
wt	0,89	0,89

G228N + T230S	G/(G+NG)
	slow
A ^{-/-}	0,84
B ^{-/-}	0,61

G228N + T230S	G/(G+NG)
	fast
A ^{-/-}	0,84
B ^{-/-}	0,63

Tabla 17. Cuantificación del western blot de la figura 39.

En otra transfección de las diferentes líneas HEK293 con la mutante G228N + T230S (**figura 40**), pudimos observar que en ausencia de la subunidad STT3A no hay cambios significativos con respecto a la línea wt (**tabla 18**).



Figura 40. Western blot realizado con lisados celulares de células HEK293 transfectadas con la mutante G228N+T230S.

G228N+ T230S	G/(G+NG)	
	slow	fast
wt	0,78	0,76
A ^{-/-}	0,78	0,75
B ^{-/-}	0,64	0,56

Tabla 18. Cuantificación del western blot de la figura 40.

Al transfectar a la mutante Y151S durante 24 hs en las diferentes líneas de HEK293 y luego de realizar el western blot correspondiente pudo observarse que, independientemente de la identidad celular, la proteína se encuentra 100% glicosilada:

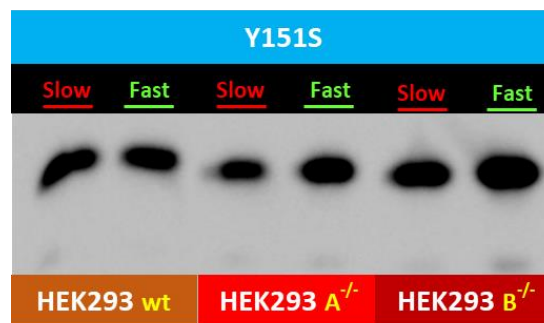


Figura 41. Western blot realizado con lisados celulares de células HEK293 transfectadas con la mutante Y151S.

Por último, para evaluar qué sucedía con un secuón que estuviera cerca del N-terminal, se generaron dos mutantes nuevas (ver figura 18):

- L7N (CTA7AAC)
- L7N + T9S (CTA7AAT+ACG9AGC)

La diferencia entre las mismas radica en la tercera posición del secuón, siendo el primero del tipo N-X-T y el segundo, N-X-S (figura 42). El propósito de tener a un mismo sitio donde sólo cambia el hidroxiaminoácido era ver si podíamos observar lo mismo que en un trabajo de Malaby y Kobertz⁸⁰, donde un sitio cercano al extremo amino terminal mostraba un alto porcentaje de modificación co-traduccional cuando el secuón poseía treonina en la posición +2, mientras que el mismo disminuía notablemente si en esa posición había serina. Usando las líneas KO, donde en teoría las B^{-/-} sólo podrían modificar al secuón co-traduccionalmente y las A^{-/-} sólo post-traduccionalmente, esperábamos poder estudiar en mayor profundidad el fenómeno ó incluso, determinar si los roles de las subunidades catalíticas de la OST estaban realmente divididos.

MDSKGSSQKGSRLLLLVSNLLCQGVVSTPLESKGEE **L** **T** GVPILVELDGDV
 NGHKFSVRGEGEGDATNGKLTCLKFISTTGKLPVPWPTLVTTLYGVQMFSRYPD
 HMKQHDFKSAPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFK
 EDGNILGHKLEYNFNHNVYITADKQKNGIKANFKIRHNVEDGS **VQLADHYQQ**
NTPIGDGPVLLPDNHVLTQSVLSKDPNEKRDHMLLEFVTAAGITHGMDELYKE
 FYPYDVPDYAGYPYDVPDYAGSYDVPDYAAKDEL

Figura 42. Ubicación del seción NFT ó NFS a generar.

Las mutantes fueron transfectadas y los resultados pueden observarse en la **figura 43**.

Dado que todas las construcciones en todas las líneas celulares dieron un 100 % de N-glicosilación, no se puede discutir nada al respecto de la ubicación e identidad del hidroxiaminoácido.

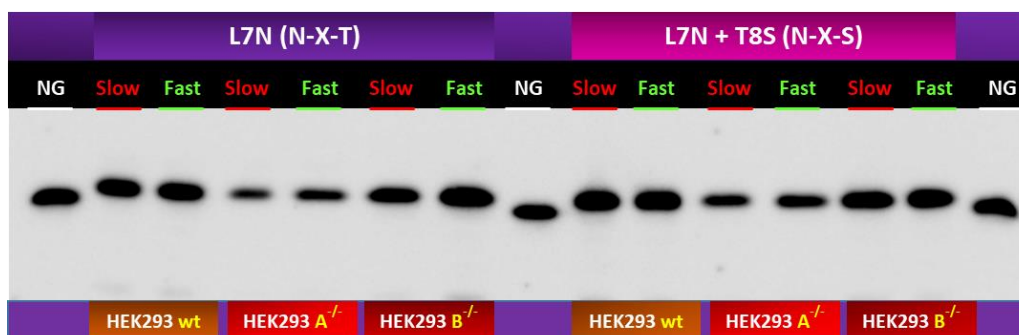


Figura 43. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

ENSAYOS DE VELOCIDAD DE SÍNTESIS PROTEICA.

Para evaluar si el criterio utilizado como medida de la velocidad de traducción fue correcto, se realizaron experimentos de marcación con ³⁵S. Los resultados se muestran en las **figuras 44 y 45**. De las ecuaciones de la recta, se calculó el tiempo mínimo a partir del cual comenzaría a haber proteína (y=0), siendo 2,8 segundos para la versión rápida de la *cf-sfGFP* y 200 segundos para la lenta (**tabla 19**). Estos resultados no resultan coherentes, al menos para la versión rápida, ya que ello significaría que una proteína de 300 aminoácidos estaría siendo sintetizada a más de 100 residuos por segundo y se ha estimado que en eucariotas la velocidad de síntesis proteica no supera los 6 aminoácidos/seg.

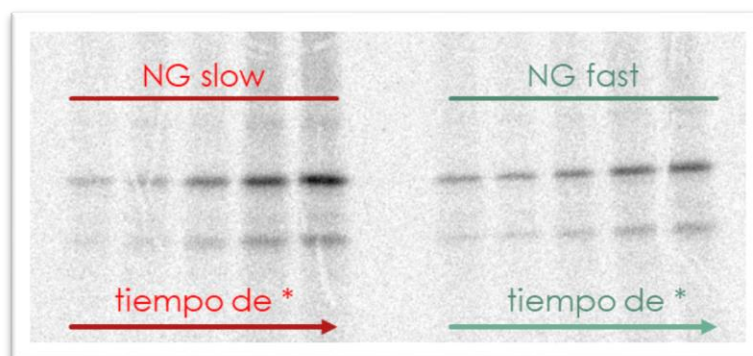


Figura 44. Resultados del gel de poliacrilamida con las proteínas marcadas con ^{35}S , revelado en el Storm.

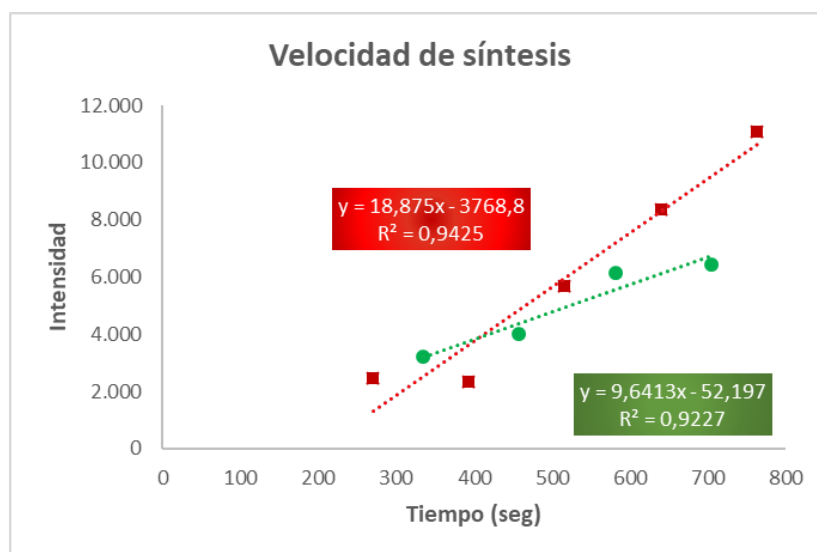


Figura 45. Cuantificación del gel de la figura 44.

Tiempo de síntesis (seg)	
slow	fast
200	2,77

Tabla 19. Tiempo de síntesis proteica calculado a partir de la ecuación de la recta de la figura 43.

El ensayo fue repetido en varias ocasiones, encontrándonos con imágenes difíciles de cuantificar debido al fondo ó con resultados disímiles. Mientras que de la tabla anterior se desprende una velocidad de traducción de 1,6 aa/seg, en la **tabla 20** se calcula una velocidad de 9,5 aa/seg para la proteína slow a partir de la **figura 46**. En el caso de la variante rápida, se sigue obteniendo un valor muy bajo para ser real (3,4 segundos).

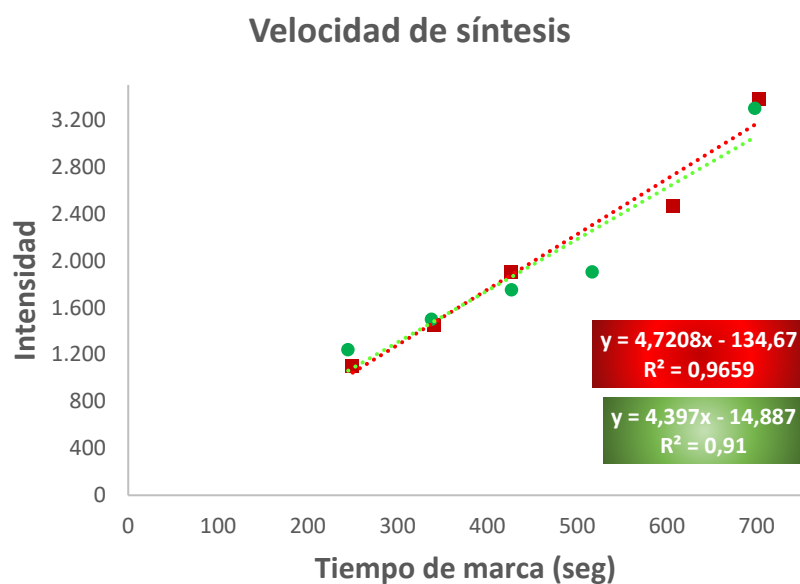


Figura 46. Gráfica resultante luego de la cuantificación del gel revelado con el Storm.

Tiempo de síntesis (seg)	
slow	fast
28,57	3,39

Tabla 20. Tiempo de síntesis proteica calculado a partir de la ecuación de la recta de la figura 46.

En un tercer ensayo (**figura 47**), se calculó una velocidad de síntesis de la variante lenta de 49 segundos (**tabla 21**), lo cual se corresponde a una traducción de 6 aa/seg. No se calcula para la variante rápida porque hubo un error experimental.

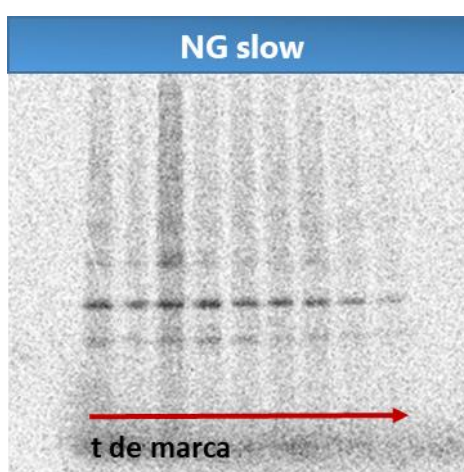


Figura 47. Resultados del gel de poliacrilamida con las proteínas marcadas con ^{35}S , revelado en el Storm.

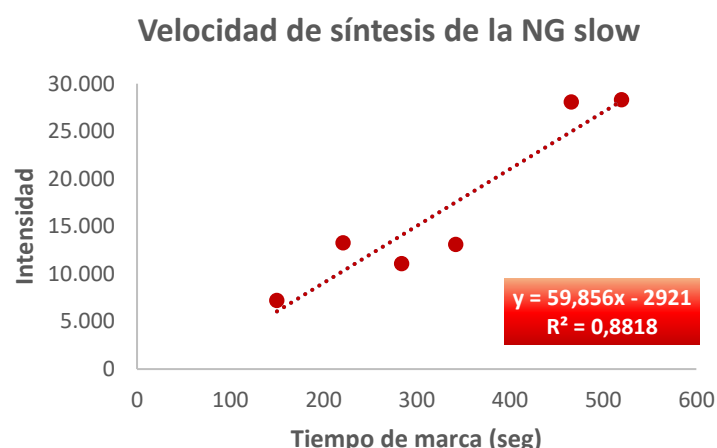


Figura 48. Cuantificación del gel de la figura 47.

Tiempo de síntesis (seg)	
slow	fast
48,8	ND

Tabla 21. Tiempo de síntesis proteica calculado a partir de la ecuación de la recta de la figura 48.

EL RETORNO DE LA C70

Una de las mutaciones iniciales en la proteína de trabajo había sido el cambio de la cisteína por metionina en la posición 70 (C70M). Nos preguntamos si ese cambio podría haber generado alguna perturbación estructural que afectara a la eficiencia del proceso de *N*-glicosilación. Entonces, decidimos revertir el cambio en varias de las construcciones ya ensayadas, para evaluar si la presencia de la cisteína 70 modificaba la ocupación del sitio de *N*-glicosilación en el estado estacionario.

Las versiones con cisteína 70 fueron generadas únicamente sobre las versiones fast. Una vez obtenidas las construcciones C70, fueron transfectadas junto con las versiones fast con metionina 70 en las diferentes líneas de HEK293 y, luego de 24 hs se procedió a la lisis celular. Con los extractos obtenidos se realizaron los western blots correspondientes, como se puede observar a continuación.

En la **figura 49** se muestran los resultados obtenidos para secuencias que están ubicados a dos residuos de distancia en la cadena polipeptídica. A pesar de ser mínimo el espaciado entre los mismos, se puede observar que claramente difieren en su comportamiento. Las mutantes NISA y NISE poseen una *N*-glicosilación notoriamente menor cuando son expresadas en células

A^{-/-} y poseen cisteína en la posición 70 (**tabla 22**). En el caso de la mutante Y151S no existen diferencias entre poseer cisteína o metionina, y la ocupación del sitio resulta máxima independientemente de la identidad del aminoácido 70.

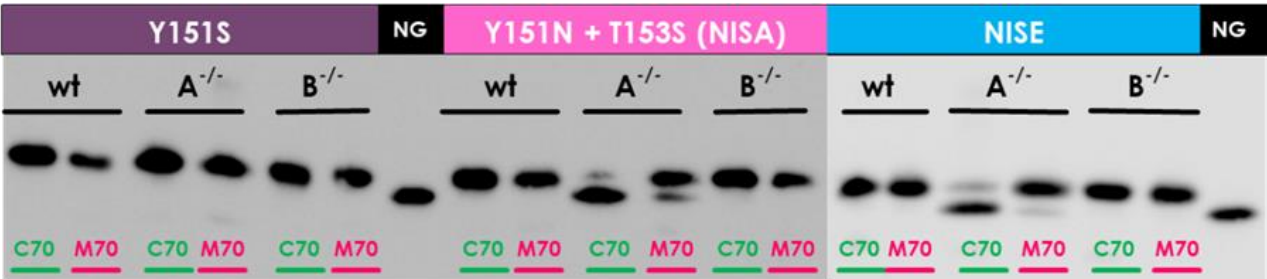


Figura 49. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

	G/(G+NG)	
	C70	M70
A ^{-/-}		
NISA	0,10	0,77
NISE	0,27	0,88

Tabla22. Cuantificación de la figura 49.

En el caso de la mutante E172S (**figura 50**), luego de cuantificar las bandas proteicas (**tabla 23**) se observa que en las células silvestres, donde la cantidad de proteína total resultó comparable entre C70 y M70, la mutante con metionina nuevamente es la que presenta un mayor grado de *N*-glicosilación.

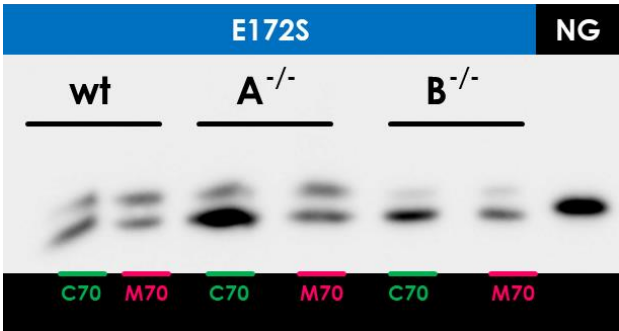


Figura 50. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

	G/(G+NG)	
	C70	M70
E172S		
wt	0,34	0,54

Tabla 23. Cuantificación de la figura 50.

Para la mutante N146S (**figura 51**) se observa un 53% de aumento en la eficiencia de *N*-glicosilación en la mutante con metionina respecto de la que tiene cisteína en las células que no poseen la subunidad STT3A (**tabla 24**). En las líneas wt y B^{-/-} no se compara porque los totales proteicos difieren entre sí.

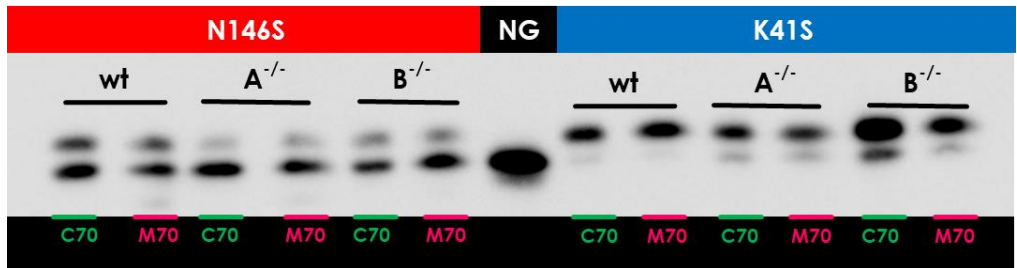


Figura 51. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

N146S	G/(G+NG)	
	C70	M70
A ^{-/-}	0,15	0,23

Tabla 24. Cuantificación del western blot de la figura 50.

En cambio, la mutante K41S (**figura 51**) parece no modificar su patrón de *N*-glicosilación al menos en las células wt y A^{-/-} (**tabla 25**).

K41S	G/(G+NG)	
	C70	M70
wt	0,90	0,93
A ^{-/-}	0,81	0,80

Tabla 25. Cuantificación del western blot de la figura 51.

En el caso de la G228N+T230S (**figura 52**) se puede ver cualitativamente que, en todas las líneas celulares, hay un menor grado de *N*-glicosilación cuando en la secuencia aminoacídica hay una cisteína 70.

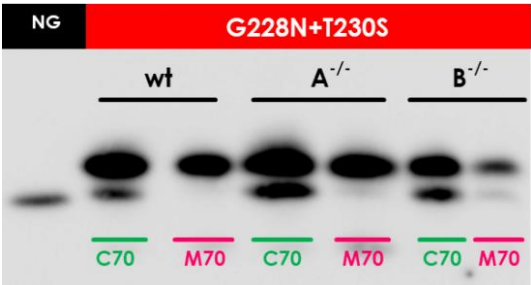


Figura 52. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas.

A excepción de la mutante Y151S, las otras ensayadas muestran que cuando poseen una cisteína en lugar de metionina en la posición 70, la eficiencia de *N*-glicosilación se ve reducida. Ante dichos resultados, nos preguntamos si esa disminución estaría relacionada con la posible formación de puentes disulfuro entre proteínas con cisteína 70 y, para dilucidar el asunto decidimos hacer un gel en presencia/ausencia de agente reductor (β -mercaptoetanol, **figura 53**).

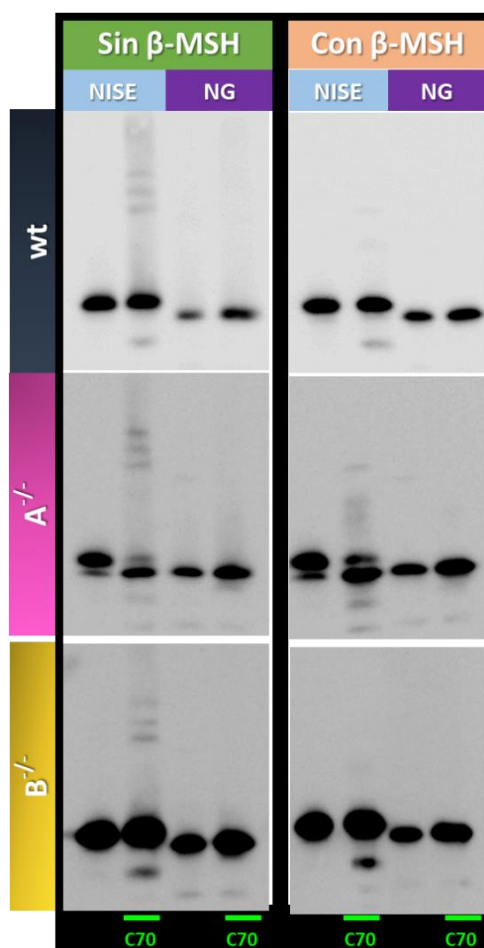


Figura 53. Western blot realizado con lisados celulares de células HEK293 transfectadas con las mutantes indicadas, en presencia (izquierda) ó ausencia (derecha) de agente reductor.

En la **figura 53** se puede ver que, en el caso de la NISE, existe una muy baja tendencia a la oligomerización covalente, y que la misma no puede dar cuenta de los cambios que se muestran en las células que carecen de la subunidad STT3A, ya que esa mínima formación se encuentra presente también en las células wild type y en las no poseen a la subunidad STT3B. Por otra parte, la C70 se encuentra oculta del solvente dentro del barril beta de la *sfGFP* y no es de esperar que forme puentes disulfuro una vez que la proteína se haya plegado.

Descartado que el fenómeno se debiera a la formación de agregados, nos preguntamos si podría ser una consecuencia de que las variantes **C70** se estuviesen plegando más rápido, estando menos tiempo el secunón accesible a la OST o, alternatively, que estuviera aumentada la estabilidad termodinámica de esta proteína respecto de la mutante **C70M**.

EXPRESIÓN Y PURIFICACIÓN DE LAS *sfGFPs*

Para comprender las diferencias observadas en la *N*-glicosilación al introducir la mutación **C70M** se procedió a expresar de forma recombinante en *E. coli* a las versiones no glicosiladas de estas proteínas. Para ello se clonaron en el vector pET22b las secuencias codificantes de las versiones maduras de las mismas, desprovistas de péptido señal, el epítipo 3xHA y la señal de retención en el RE. La expresión y purificación de las proteínas fue analizada mediante electroforesis en SDS PAGE (**figura 54**). Ambas proteínas migran de acuerdo a su masa molecular esperada (27,9 kDa). Las dos proteínas se encontraron mayoritariamente en la fracción soluble del lisado celular (comparar calles 2 y 3 para **C70** y calles 7 y 8 para **M70**), indicando que se encuentran plegadas correctamente. Luego del paso de purificación mediante IMAC se observan algunas impurezas de menor peso molecular (calles 4 y 9), las cuales fueron eliminadas mediante la cromatografía de exclusión molecular (calles 5 y 10). El rendimiento de las purificaciones fue muy bueno en ambos casos (28 y 30 mg/l de cultivo para **C70** y **M70**, respectivamente).

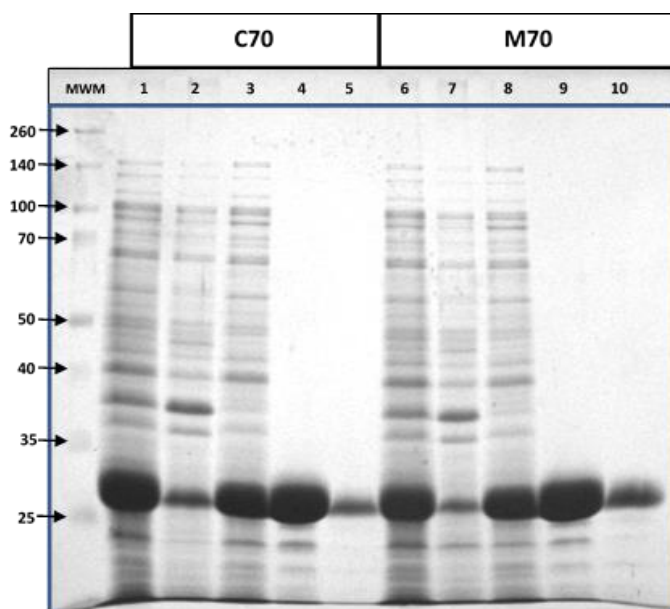


Figura 54. Expresión y purificación de las *sfGFPs* C70 y M70. 1 y 6: lisado total de bacterias. 2 y 7: fracción insoluble. 3 y 8: fracción soluble. 4 y 9: eluido de la IMAC. 5 y 10: eluido de la Superdex 200. **MWM**: marcador de masa molecular (Thermo Spectra BR).

El estado de agregación de las proteínas fue analizado mediante cromatografía de exclusión molecular siguiendo la elución por absorbancia a 280 nm. En ambos casos se observó un pico mayoritario que eluye a 16.0 ml (**figura 55A**), viéndose también un ligero hombro que eluye antes. El mismo podría deberse a la formación de dímeros de las proteínas o la presencia de alguna impureza de mayor peso molecular. Para comparar el perfil de ambas proteínas se superpusieron los perfiles normalizados de ambos cromatogramas. Se puede observar que los perfiles son idénticos (**figura 55B**).

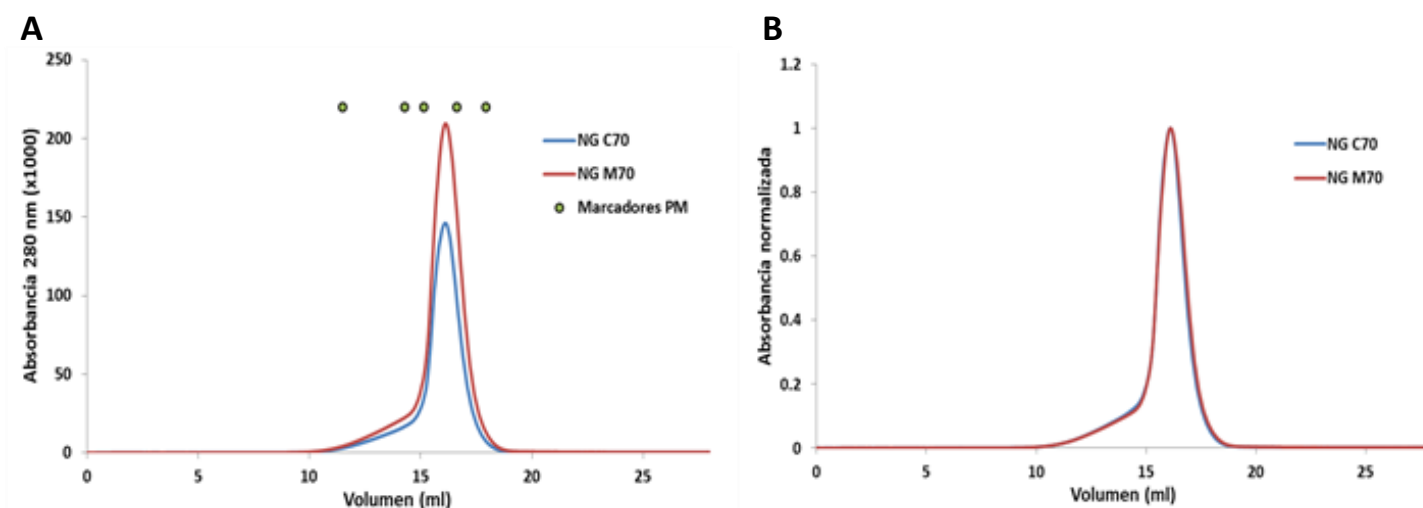


Figura 55. Análisis por cromatografía de exclusión molecular de las proteínas recombinantes. (A) Perfil de elución de las proteínas *sfGFPs* C70 y M70 seguido por absorbancia a 280 nm. Los círculos indican el volumen de elución de marcadores de peso molecular. **(B)** Perfiles normalizados por la absorbancia máxima.

A partir del volumen de elución se calculó el PM de ambas proteínas. Para ello se calibró la columna con proteínas de peso molecular conocido (Aldolasa (158 kDa), Conalbúmina (75 kDa), Ovalbúmina (43 kDa), Anhidrasa carbónica (23 kDa) y Ribonucleasa A (13.7 kDa)). El peso medido para ambas variantes fue de 30.4 kDa, muy cercano al peso teórico calculado a partir de su composición de aminoácidos (**figura 56**). Estos resultados indican que las proteínas se encuentran en forma monomérica.

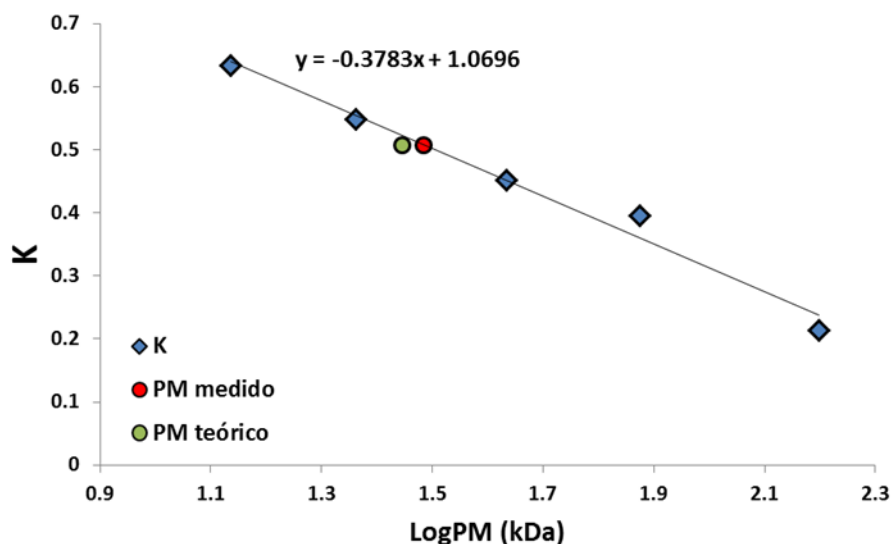


Figura 56. Estimación del peso molecular de las sfGFPs. Los rombos azules indican la K de retención para los estándares de peso molecular. La misma fue calculada como $K = (V_e - V_0)/(V_0 - V_t)$, en donde V_e es el volumen de elución, V_0 es el volumen de exclusión de la columna y V_t es su volumen total. Los puntos verde y rojo indican los pesos moleculares medidos y teóricos para la sfGFP, respectivamente.

Determinación de la estructura secundaria y estabilidad térmica

La estructura secundaria de las proteínas se estudió mediante dicróismo circular en el ultravioleta lejano. Ambos espectros mostraron una banda de magnitud negativa centrada a 216 nm y una banda positiva cerca de los 200 nm (**figura 57**). Estos espectros son característicos de hoja plegada beta y concuerdan con la estructura cristalográfica de la sfGFP.

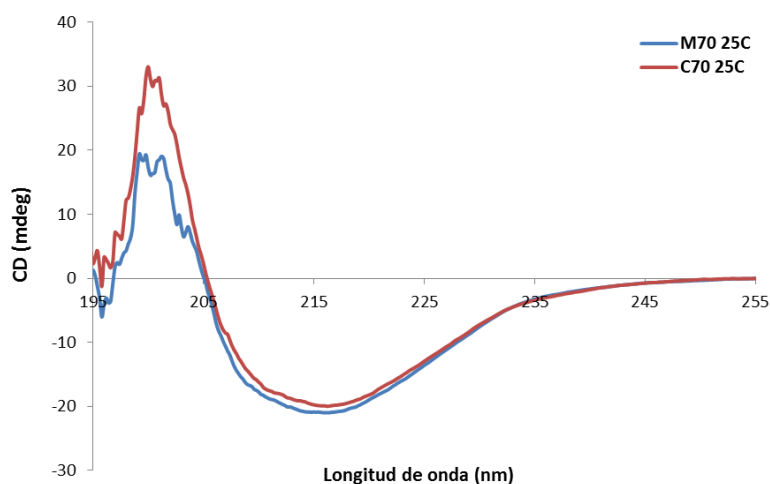


Figura 57. Espectro de dicróismo circular de sfGFP C70 y M70.

Los espectros de la **figura 57** fueron tomados a 25°C. Cuando las muestras fueron medidas a 95°C se observó un cambio drástico en su forma, indicando un proceso de

desnaturalización (**figura 58**). Para estudiar la reversibilidad de este proceso, las proteínas calentadas a 95°C fueron llevadas nuevamente a 25°C y luego de 10 minutos se volvió a medir el espectro. Se puede observar que los espectros son similares a los medidos antes del calentamiento, indicando que el desplegado térmico es en gran medida reversible. Este comportamiento es infrecuente en proteínas de este tamaño, puesto que usualmente la agregación proteica durante el enfriamiento disminuye la eficiencia de renaturalización.

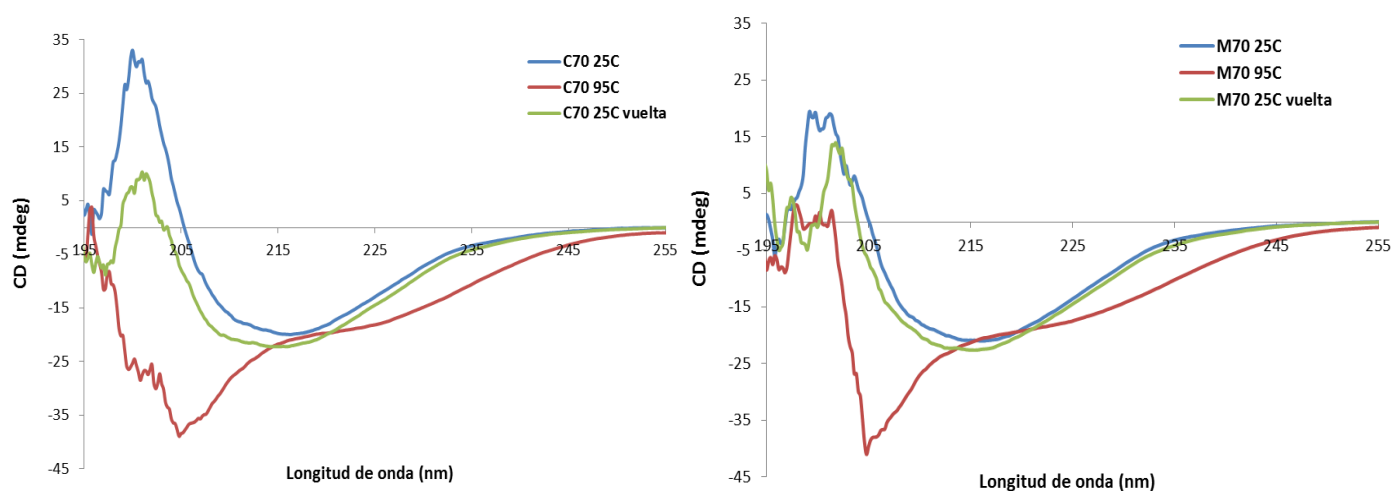


Figura 58. Efecto de la temperatura sobre los espectros de CD. (A) C70. (B) M70. “C70 25C vuelta” y “M70 25C vuelta” fueron medidos luego de volver a la temperatura inicial de 25°C.

La temperatura de fusión (T_m) de las proteínas se determinó siguiendo el cambio en la señal de CD a 232 nm. Esta longitud de onda presenta una buena relación entre el ruido y el cambio de señal entre el estado nativo y el desplegado. Un modelo de equilibrio a dos estados fue ajustado a los datos experimentales. Dicho modelo tiene en cuenta la variación en los parámetros termodinámicos en función de la temperatura, los cuales se relacionan con el cambio de capacidad calorífica entre los estados nativo ($C_{p,n}$) y desplegado ($C_{p,u}$):

$$\Delta C_p = C_{p,u} - C_{p,n}$$

Si consideramos que ΔC_p es independiente de la temperatura en el rango empleado, se puede expresar ΔH y ΔS como:

$$\Delta H(T) = \Delta H(T_0) + \Delta C_p (T - T_0)$$

$$\Delta S(T) = \Delta S(T_0) + \Delta C_p \ln (T / T_0)$$

Con estas ecuaciones se calcula el cambio en la energía libre de desplegado en función de la temperatura:

$$\Delta G_u(T) = \Delta H(T_0) + \Delta C_p (T - T_0) - T \cdot (\Delta S(T_0) + \Delta C_p \ln (T / T_0))$$

De aquí se puede calcular la dependencia de la constante de desplegado con la temperatura. En este tipo de curvas es frecuente observar pendientes diferentes de cero en las señales de las pre- y post-transiciones. Este efecto es tenido en cuenta asumiendo una dependencia lineal con la temperatura de la señal de los estados nativo y desplegado. En la temperatura de fusión $\Delta G_u = 0$ y se cumple que $\Delta H_u(T_m) = T_m \cdot \Delta S_u(T_m)$. Si tomamos como T_0 de referencia a T_m :

$$\Delta G_u(T) = \Delta H(T_m) (1 - T / T_m) + \Delta C_p (T - T_m - T \cdot \ln (T / T_m))$$

Siendo la dependencia de la intensidad de la señal con T:

$$I(T) = (I_{n,0} + b_n \cdot T) + \frac{\exp (-(\Delta H(T_m) (1 - T/T_m) + \Delta C_p (T - T_m - T \cdot \ln (T / T_m)))/RT)}{1 + \exp (-(\Delta H(T_m) (1 - T/T_m) + \Delta C_p (T - T_m - T \cdot \ln (T / T_m)))/RT)} (I_{u,0} - I_{n,0} + T \cdot (b_u - b_n))$$

En donde $I_{n,0}$ e $I_{u,0}$ son las señales de de los estados N y U a cero Kelvin, respectivamente, mientras que b_n y b_u son las dependencias de estas señales con la temperatura. Este modelo se pudo ajustar de forma satisfactoria a los datos experimentales. Las curvas de desnaturalización fueron medidas en buffer o en presencia de urea 2 M. Esto último se realizó para poder determinar de forma más precisa la T_m de la variante **C70**, puesto que su alta estabilidad no permite medir bien la post-transición en ausencia de urea (**figura 59**).

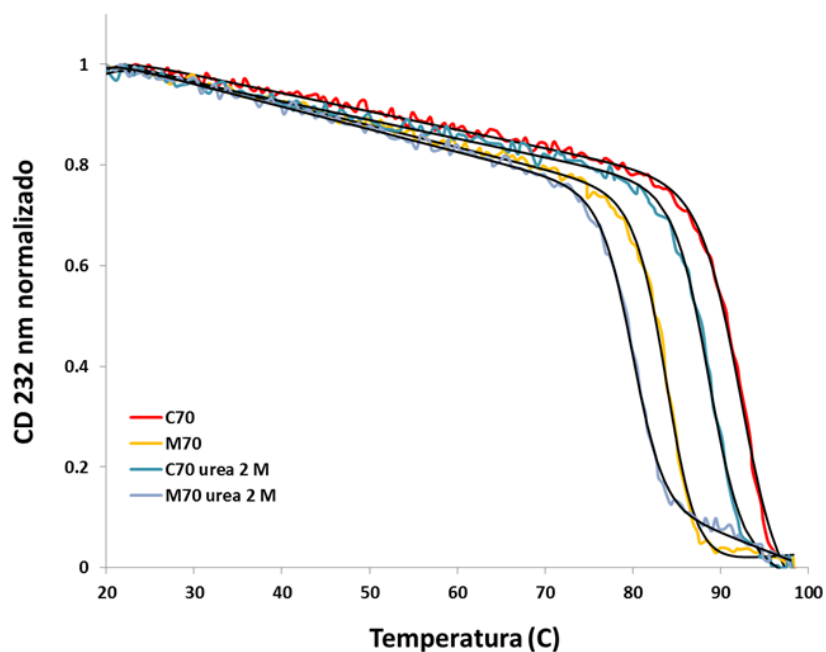


Figura 59. Curvas de desplegamiento térmico de sfGFP C70 y M70 seguidas por CD. La señal de CD a 232 nm fue normalizada para cada curva.

A partir del modelo ajustado se pudo calcular la **T_m** de ambas proteínas en las dos condiciones experimentales (**Tabla 26**).

	T_m urea 0 M	T_m urea 2 M	Delta T_m
C70	92.0 ± 0.4 °C	88.4 ± 0.3 °C	3.6 °C
M70	83.8 ± 0.3 °C	79.8 ± 0.2 °C	4.0 °C
Delta T_m	8.2 °C	8.6 °C	

Tabla 26. Temperaturas de desplegamiento de las sfGFP.

Se puede observar que la urea indujo la desnaturalización de las proteínas a menor temperatura, siendo el efecto similar para ambas (**Tabla 26**, cuarta columna). Este comportamiento es esperable dado el efecto desestabilizante de la urea. Es destacable la enorme estabilidad térmica de ambas proteínas. La variante **C70** es mucho más estable, presentando una T_m 8.2°C mayor a 0 M urea (**Tabla 26**, cuarta fila).

Estabilidad conformacional frente a cloruro de guanidinio

La gran diferencia observada en la estabilidad térmica de las proteínas nos llevó a estudiar este fenómeno en más detalle. Para ello se realizaron curvas de desplegamiento químico inducido por cloruro de guanidinio (GdnHCl). Esta metodología permite medir el cambio de energía libre (**ΔG_u**) del desplegamiento. Las muestras fueron incubadas a 25°C con concentraciones

creciente de GdnHCl durante 96 hs. Dicho tiempo es necesario para asegurar que el sistema haya alcanzado el equilibrio. Las curvas de desplegamiento fueron seguidas mediante espectroscopía de fluorescencia. Se midieron dos series de espectros, el espectro del cromóforo natural de las *sfGFP* empleando una longitud de onda de excitación de 498 nm y el espectro del único residuo de triptófano que presentan estas proteínas (**W57**) usando una longitud de onda excitación de 295 nm. Esta metodología permite seguir el proceso de desplegamiento mediante dos reporteros diferentes, pudiendo evidenciar la presencia de intermediarios. La fluorescencia del cromóforo fue totalmente apagada al desnaturizarse las proteínas (**figura 60 A y B**). En cuanto a la fluorescencia del triptófano, se pudo observar un corrimiento marcado hacia el rojo del espectro con un aumento en la intensidad de la señal (**figura 60 C y D**).

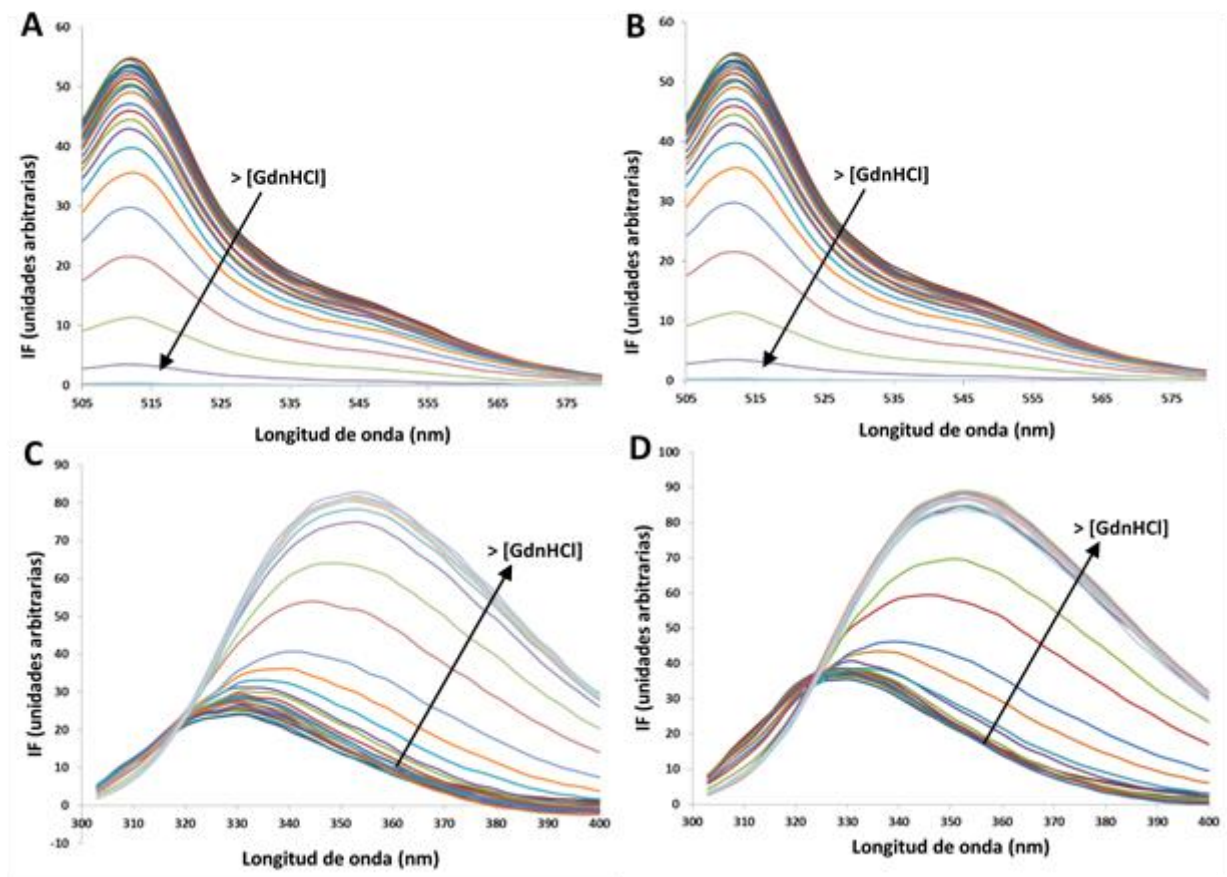


Figura 60. Desplegado inducido por GdnHCl. El desplegamiento fue seguido mediante la fluorescencia del cromóforo de las *sfGFPs* (paneles **A** y **B**) o del residuo de triptófano (paneles **C** y **D**). Variante **C70**: paneles **A** y **C**. Variante **M70**: paneles **B** y **D**.

El máximo de fluorescencia de triptófano en ausencia de desnaturizante se encuentra a 330 nm, valor característico de un entorno no polar. Dicho máximo se mueve a 353 nm luego del desplegamiento, indicando su exposición al solvente. Por otra parte, el aumento de intensidad

en el desplegado puede explicarse por un efecto de apagamiento que sufre el triptófano en el estado plegado de la proteína. La estructura cristalográfica de la variante **C70** muestra al triptófano en contacto con el residuo de metionina M218 (**figura 61**). Las metioninas apagan eficientemente la fluorescencia de los triptófanos que se encuentran en contacto. La pérdida de este contacto durante el desplegado explicaría entonces el aumento en la intensidad de los espectros.

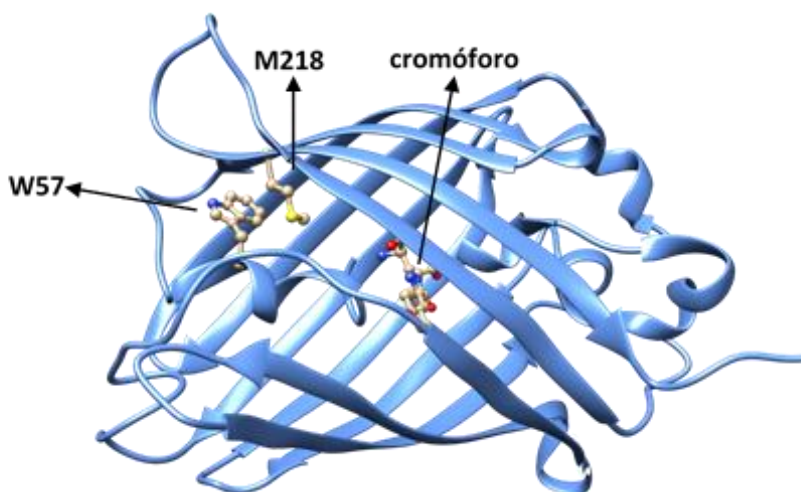


Figura 61. Disposición espacial de los residuos W57 y M218 en la estructura de la sfGFP C70. La figura fue realizada con el programa Chimera a partir del archivo PDB **2B3P**.

El desplegado fue analizado mediante un modelo de equilibrio entre dos estados. El cambio de energía libre del desplegado ΔG_u depende linealmente con la concentración de desnaturante **[D]**:

$$\Delta G_u (D) = \Delta G_{u,o} - m [D]$$

En donde $\Delta G_{u,o}$ es el cambio de energía libre en ausencia de desnaturante, y **m** da cuenta de la cooperatividad de la transición. Usualmente este parámetro está relacionado al cambio de superficie accesible al solvente durante el desplegado. La intensidad de fluorescencia en función de la concentración de desnaturante puede describirse mediante la siguiente ecuación:

$$I(D) = (I_{n,o} + b_n \cdot [D]) + \frac{\exp (-(\Delta G_{u,o} - m [D])/RT)}{1 + \exp (-(\Delta G_{u,o} - m [D])/RT)} (I_{u,o} - I_{n,o} + [D] \cdot (b_u - b_n))$$

En donde $I_{n,0}$ e $I_{u,0}$ son las señales de los estados N y U en ausencia de desnaturalizante, respectivamente, mientras que b_n y b_u son las dependencias de estas señales con la concentración de desnaturalizante. Este modelo se pudo ajustar de forma satisfactoria a los datos experimentales (figura 62).

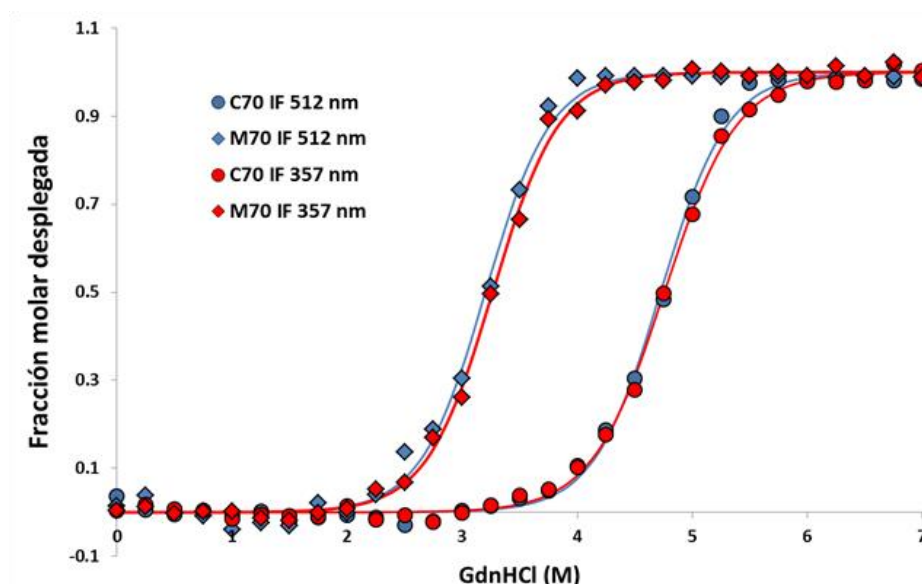


Figura 62. Desplegado de sfGFPs inducido por GdnHCl seguido por fluorescencia. La fluorescencia a 357 (símbolos rojos) y 512 nm (símbolos azules) para las variantes **C70** (círculos) y **M70** (rombos) en función de la concentración de GdnHCl. Las curvas representan el resultado del ajuste a un modelo de dos estados para cada señal en forma independiente.

A partir del ajuste de un modelo de dos estados en equilibrio se pueden calcular los parámetros termodinámicos de la transición (tabla 27). Los cambios de señal normalizados medidos a 512 nm y a 357 nm son muy similares, indicando la ausencia de intermediarios de plegado detectables. Esto valida el uso de un modelo de dos estados para describir el proceso.

	C70 512 nm	C70 357 nm	M70 512 nm	M70 357 nm
$\Delta G_{u,0}$ (kJ/mol)	40.3 ± 1.0	37.3 ± 0.9	28.1 ± 1.0	28.0 ± 1.0
m (kJ/mol.M)	8.5 ± 0.3	8.8 ± 0.2	7.8 ± 0.2	8.5 ± 0.3
GdnHCl 50 % (M)	4.7	4.8	3.2	3.3

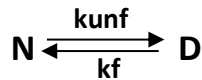
Tabla 27. Parámetros termodinámicos calculados a partir de la figura 62.

De forma similar a lo observado en CD, las curvas de fluorescencia mostraron que la mutación **C70M** genera una marcada desestabilización de la proteína. La disminución en el cambio energía libre de desplegado es cercana a los 12 kJ/mol (medido a 512 nm). Por otra parte, en todos los casos el parámetro m es similar, indicando un cambio similar de área

accesible al solvente durante el desplegado. El efecto desestabilizador de la mutación puede deberse a un estrés conformacional generado por el mayor volumen de la metionina respecto de la cisteína. Si bien los espectros de fluorescencia y CD en ausencia de desnaturante son muy similares para ambas proteínas, claramente la mutación perjudica la estabilidad de la proteína.

Cinética de plegado y desplegado

Desde un punto de vista cinético, la menor estabilidad de la variante **M70** puede deberse a que una menor velocidad de plegado y/o a una mayor velocidad de desplegado. En el caso más simple, el proceso de plegado/desplegado puede describirse como dos reacciones de primer orden:



A medida que aumenta la concentración de agente desnaturante, la velocidad de desplegado sube y la de plegado disminuye. En la concentración en que ambos procesos proceden a la misma velocidad, luego de que el sistema llega al equilibrio, se tiene que $[N] = [D]$, $k_{unf} = k_f$ y $\Delta G_u = 0$. Este punto se corresponde con la mitad de la transición de la curva de desplegado medida al equilibrio. En los casos más simples se observa que el logaritmo de las constantes cinéticas varía linealmente con la concentración de desnaturante:

$$\ln(k_{unf}) = \ln(k_{unf,0}) + m_{unf} \cdot [D]$$

$$\ln(k_f) = \ln(k_{f,0}) - m_f \cdot [D]$$

En donde m_{unf} y m_f son constantes de magnitud positiva que describen la dependencia de las constantes cinéticas respectivas con la concentración de desnaturante, y $k_{unf,0}$ y $k_{f,0}$ son las constantes en ausencia de desnaturante. Para estudiar el desplegado de las *sfGFPs* se partió de las proteínas en buffers y se las diluyó en concentraciones finales variables de GdnHCl. Para ello se empleó un mezclador rápido de Applied Photophysics y se midió en tiempo real la pérdida de fluorescencia de las proteínas a medida que se desnaturizaban. Como se esperaba, en ambos casos se observó que la velocidad de desplegado aumenta a medida que aumenta la concentración de GdnHCl (**figura 63**).

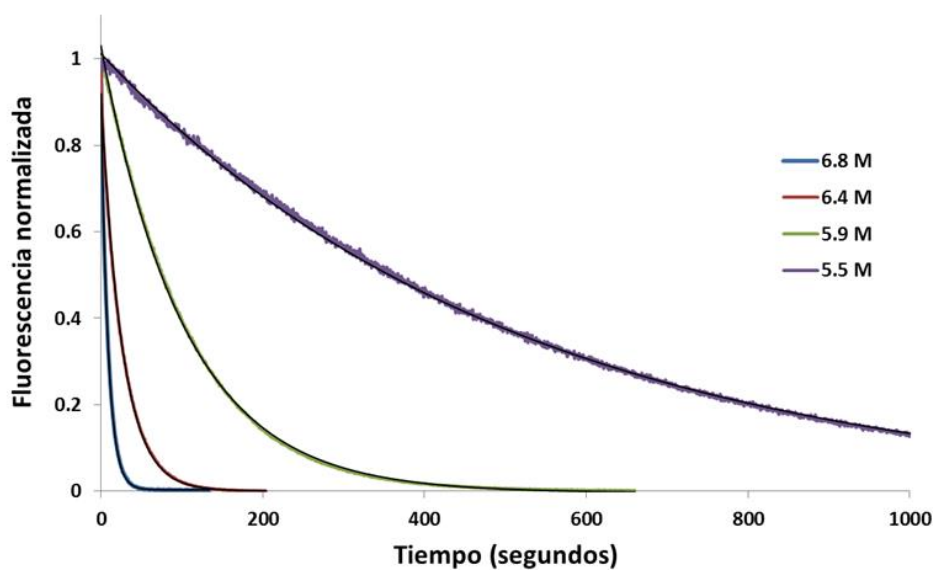


Figura 63. Desplegado de sfGFP M70 seguida por fluorescencia. Intensidad de fluorescencia normalizada durante el desplegamiento de sfGFP **M70** en distintas concentraciones de GdnHCl. Las líneas negras finas dentro de cada trazado corresponden al ajuste de un modelo monoexponencial.

Al estudiar el desplegamiento de la variante **C70** se observó que en todas las concentraciones empleadas de GdnHCl esta proteína se despliega con una cinética mucho más lenta (**figura 64**).

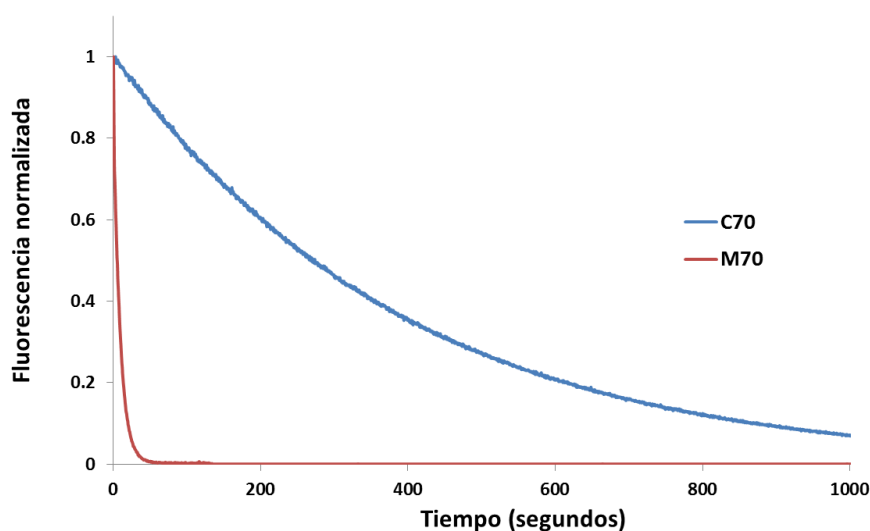


Figura 64. Comparación de la cinética de desplegamiento. La concentración final de GdnHCl para ambas curvas fue de 6.8 M.

Un modelo monoexponencial de la forma $F(t) = F_0 \cdot (1 - \exp(-t \cdot k))$ fue ajustado de forma satisfactoria a todas las curvas experimentales. Una mayor k indica una mayor velocidad de desplegamiento. El logaritmo natural de dichas constantes mostró una variación lineal con la concentración de GdnHCl, siendo en todos los casos las constantes de la mutante **C70M** mucho mayores (**figura 65**).

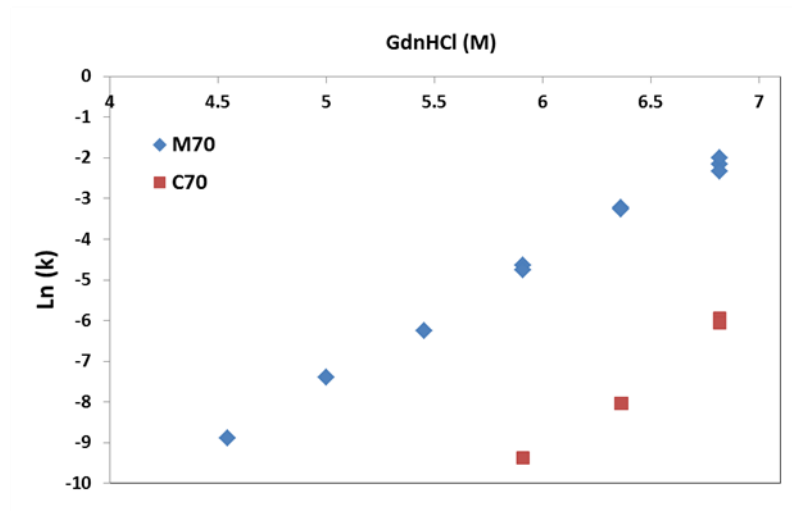


Figura 65. Variación de las constantes cinéticas medidas en la zona de desplegado en función de la concentración de GdnHCl.

La velocidad del proceso de re-plegado fue estudiada en forma similar, viéndose incrementada al disminuir la concentración de GdnHCl (**figura 66**).

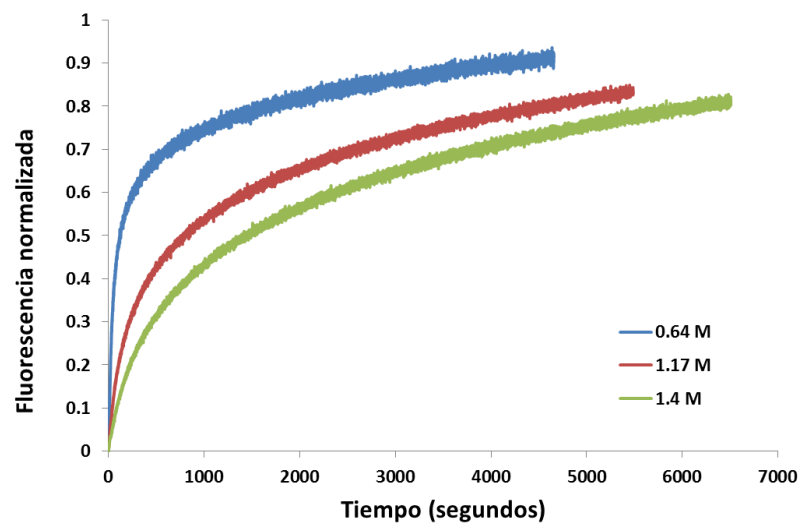


Figura 66. Cinética de replegado de sfGFP C70 en diferentes concentraciones de desnaturizante.

Contrariamente a lo observado para el desplegado, el proceso de replegado de ambas proteínas presenta una cinética muy similar para ambas proteínas (**figura 67**).

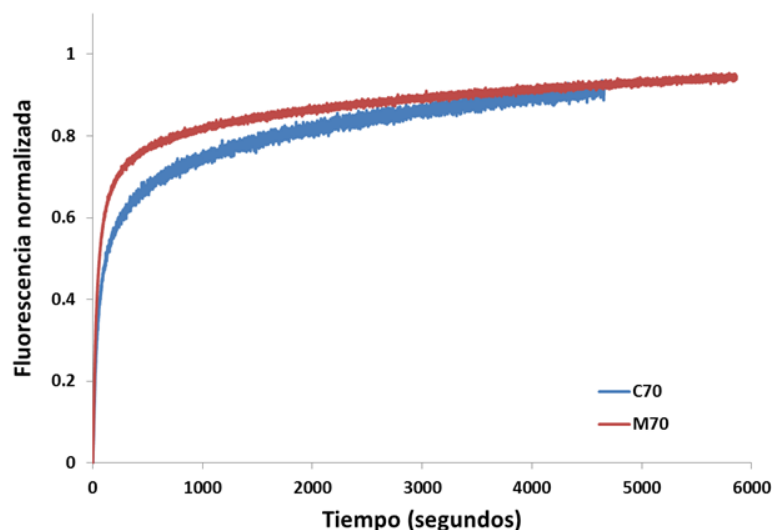


Figura 67. Cinética de replegado de las variantes C70 y M70 en 0.64 M GdnHCl.

Al intentar realizar un ajuste de un modelo mono exponencial a los datos de replegado se evidenció que este proceso sigue una cinética mucho más compleja, en donde los residuales del ajuste mostraron una tendencia no azarosa. Por tal motivo se ajustaron modelos multiexponenciales con hasta cuatro componentes independientes (figura 68).

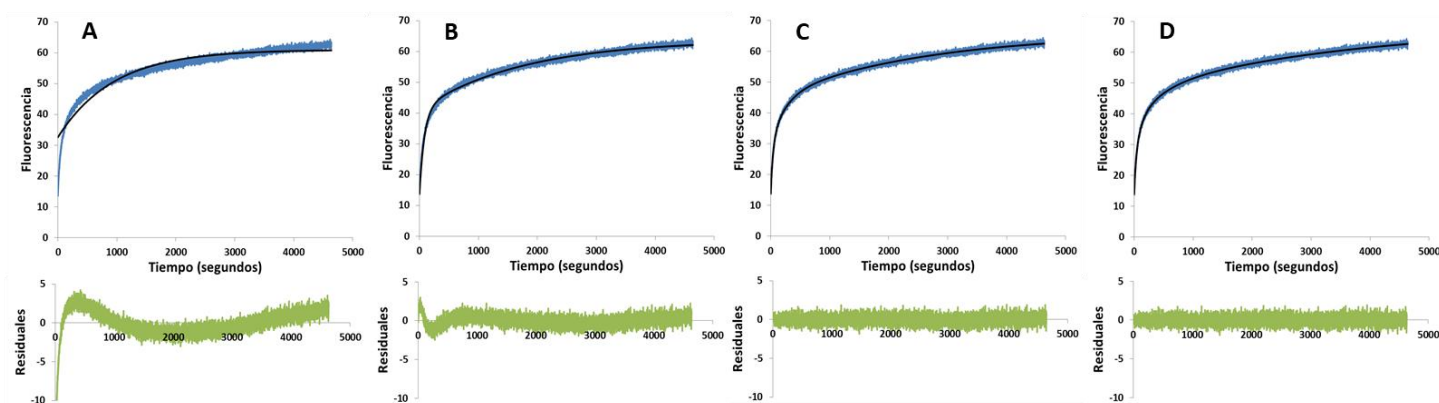


Figura 68. Ajuste de las curvas de replegado a diversos modelos exponenciales. (A) Mono-exponencial. (B) Dos exponenciales. (C) Tres exponenciales. (D) Cuatro exponenciales. Las curvas azules son los datos correspondientes al replegado de la variante **C70** en GdnHCl 0.64 M y las curvas negras son los modelos ajustados. Debajo de cada gráfico se muestran los residuales de cada ajuste.

Si bien en principio no hay límite para la cantidad de componentes exponenciales en los modelos de ajuste, se alcanza un punto a partir del cual el resultado no mejora. Para ilustrar esto se graficó la sumatoria de los desvíos cuadráticos de cada ajuste en función del número de componentes empleadas (figura 69). Se puede observar que ya con tres componentes se logra un ajuste muy bueno.

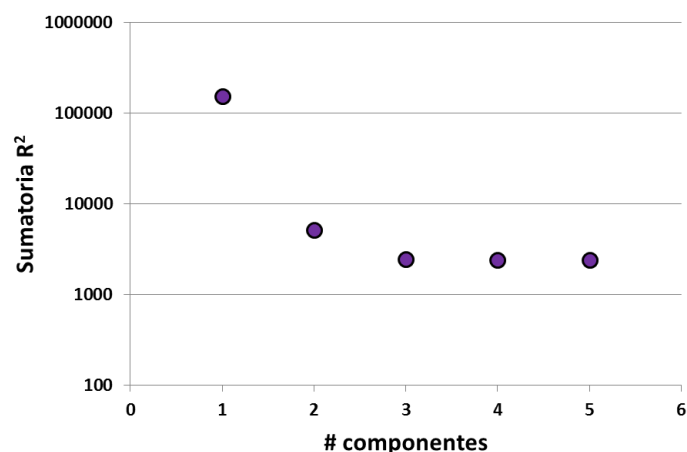


Figura 69. Variación de la sumatoria de residuales al cuadrado en función del número de componentes exponenciales empleados en el ajuste.

La complejidad del proceso de repliegado es algo esperable para una proteína como la *sfGFP*. En su estado nativo esta proteína presenta 9 prolinas en configuración *trans* y una prolina en configuración *cis* (**P89**). De hecho, este último residuo se encuentra en la vecindad del residuo mutado **C70** (figura 70).

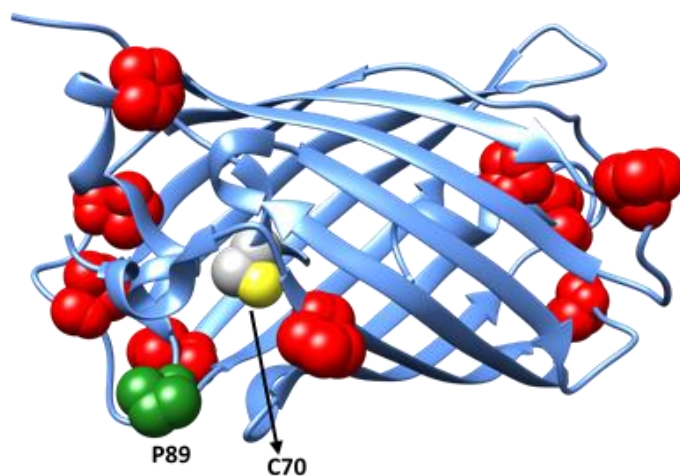


Figura 70. Ubicación de los residuos de prolina en la *sfGFP*. Las prolinas en configuración *trans* se ilustran en rojo y en verde a la **P89** en *cis*.

Cuando la proteína es incubada a alta concentración de GdnHCl los residuos de prolina están libres de las restricciones estructurales impuestas por el estado nativo y se equilibran entre conformaciones *trans* y *cis* en una proporción cercana a 95:5. Como resultado de este proceso, el estado desplegado en realidad es una mezcla de las diversas combinaciones de isomerización de las prolinas. Durante el proceso de repliegado estos residuos deben adoptar la conformación que presentan en el estado nativo. El problema radica en la alta energía de

activación de este proceso, cercano a los 100 kJ/mol, siendo en muchos casos el paso limitante del plegado. Aquellas proteínas que desplegadas necesiten isomerizar menos prolinas requerirán menos tiempo para plegarse. Muy posiblemente esto sea el origen del complejo comportamiento observado. Dada la enorme cantidad de prolinas que presenta esta proteína, con los datos disponibles no es posible discriminar los diversos caminos que tienen lugar durante el replegado.

SIMULACIONES DE DINÁMICA MOLECULAR

Puesto que el cambio de Cys por Met en la estructura de la GFP tiene efectos muy diversos en la eficiencia de ocupación de los sitios de *N*-glicosilación, decidimos estudiar ambas proteínas mediante simulaciones de dinámica molecular. El objetivo de este estudio es obtener una comprensión estructural de los cambios observados. Estos análisis fueron realizados en colaboración con el Dr. Carlos Manuel Alberto Guardia, quien actualmente se encuentra realizando su post doctorado en el NIH.

Un parámetro que permite estudiar la movilidad de la estructura a lo largo de la simulación es la raíz cuadrada de la desviación cuadrática media (RMSD, del inglés root mean squared deviation) respecto a una estructura de referencia. La misma se calcula como:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

donde *N* corresponde al número de átomos en el sistema, y δ_i a la distancia entre las posiciones del átomo *i* en dos estructuras distintas, una de ellas será la estructura de referencia. Realizando este cálculo para cada una de las “fotos” de la simulación, se obtiene el RMSD en función del tiempo. La estructura de referencia puede ser la estructura cristalográfica, por ejemplo, o la estructura promedio, lo que otorga una noción de la magnitud de las fluctuaciones del sistema a lo largo de la dinámica, es decir, como una función del tiempo. Ahora, si el promedio es sobre el tiempo, se puede tener una noción de las fluctuaciones de cada partícula:

$$RMSF = \sqrt{\frac{1}{T} \sum_{t_j=1}^T [x_i(t_j) - \bar{x}_i]^2}$$

donde T es el tiempo total de promedio y \bar{x}_i es la posición de referencia de la partícula i . Esta es la expresión para calcular la raíz cuadrada de la fluctuación cuadrática media (RMSF), la cual resulta de mucha utilidad para comparar las magnitudes de las fluctuaciones de cada partícula del sistema en un rango de tiempo de simulación determinado, es decir, como una función de cada partícula.

Al realizar la simulación para las versiones de la proteína con Cys70 ó Met70, no se encontraron cambios significativos en el RMSD luego de 20 ns (**figura 71**), por lo que cada proteína se encuentra en un mínimo de energía local y se pueden utilizar los últimos 10-20 ns para analizar las estructuras.

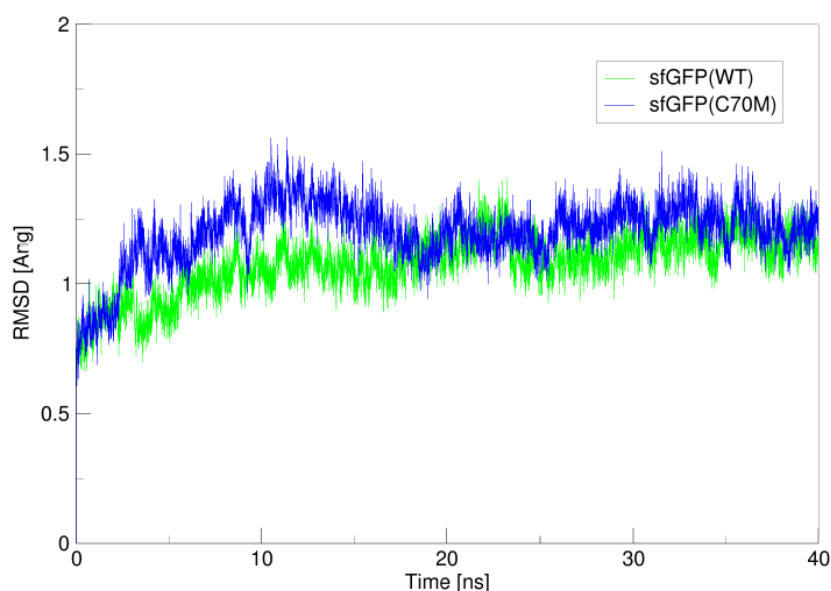


Figura 71. RMSD en función del tiempo de simulación para la *sfGFP C70* (verde) ó *sfGFP M70* (azul).

El análisis del RMSF muestra que la movilidad a nivel de los residuos individuales es similar para ambas proteínas (**figura 72A**), siendo en general mayor en los extremos N- y C-terminales y en los loops (**figura 72B**).

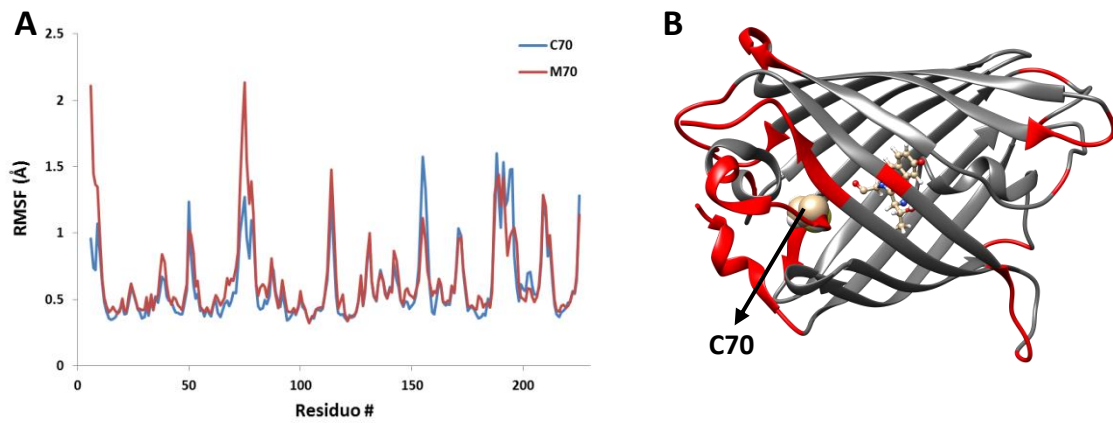


Figura 72. (A) Movilidad de los carbonos alfa a lo largo de las secuencias de las variantes C70 (línea azul) y M70 (línea roja). (B) Representación en cintas de la sfGFP en donde las regiones con movilidad de los carbonos alfa mayor a 0.7 Å fueron marcadas en rojo.

Para comparar los cambios de movilidad entre ambas proteínas se calculó la diferencia en el RMSF a lo largo de la secuencia (**figura 73A**). La introducción de la mutación **C70M** produjo cambios en cuatro regiones. Mientras que en el extremo N-terminal y la región contigua en el residuo mutado se produjo un incremento de la movilidad, los loops que conectan las hebras 7-8 y 9-10 se encuentran más rígidos (**figura 73B**).

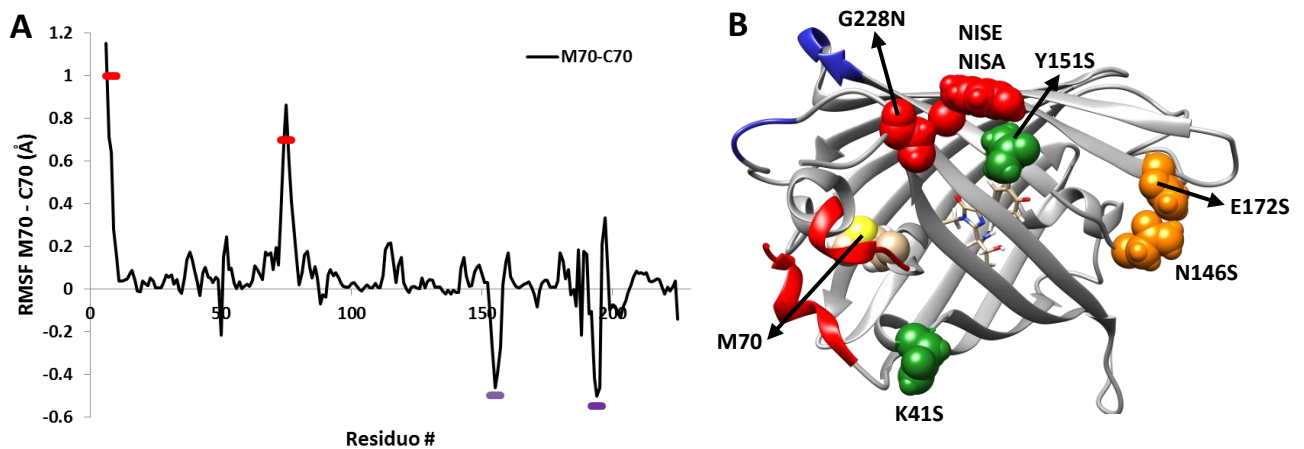


Figura 73. (A) Diferencia de movilidad de los carbonos alfa entre las variantes M70 y C70. Las barras rojas indican regiones con aumento de movilidad mayor a 0.3 Å, mientras que las barras azules indican regiones con disminución de la movilidad mayor a 0.3 Å. (B) Estructura de la variante M70 en donde se indican con color rojo y azul las zonas con aumento o disminución de movilidad, respectivamente. También se señalan los residuos en donde se introdujeron los secuencias y el residuo mutado M70.

Estos cambios fueron similares para las mutantes de glicosilación (no mostrado). En líneas generales, las posiciones en donde se ubican los secuencias no presentan grandes diferencias en su movilidad a lo largo de la dinámica entre ambas proteínas (**figura 74**). Una observación interesante es que existe una perturbación estructural en el plegado de la proteína

mutada alrededor del sitio 70 en varios loops y en el extremo C-terminal. Dichas perturbaciones se ubican cerca de los sitios que mostraron mayor cambio en el grado de glicosilación (NISE, NISA y G228N + T230S). Esto podría ser el origen del incremento en la eficiencia de *N*-glicosilación observada para estos secuencias. El resto de los secuencias se encuentran en regiones muy similares en ambas proteínas. De todas formas hay que tener en cuenta que la escala de tiempo empleada en la simulación es de sólo unas pocas decenas de nanosegundos. Un análisis de este tipo no puede mostrar fluctuaciones más amplias y menos probables que ocurren en la escala de micro a milisegundos. Por lo tanto, no puede descartarse que los cambios observados en la eficiencia de *N*-glicosilación se deban a fluctuaciones en dichas escalas de tiempo.

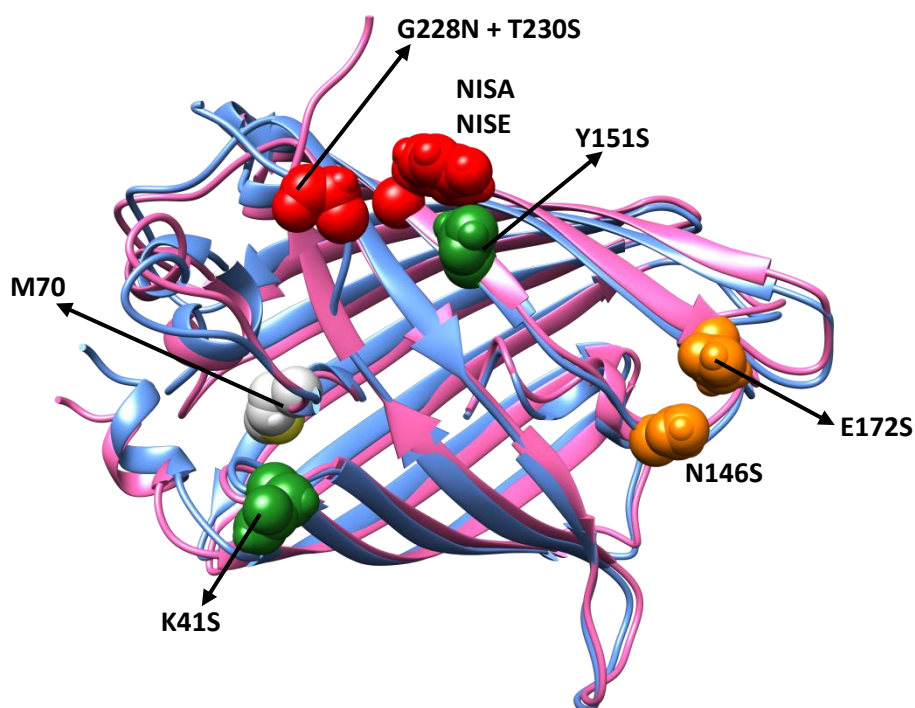


Figura 74. Reacomodamientos en la estructura de la mutante C70M. La estructura promedio a lo largo de las dinámicas de la proteína silvestre y la variante C70M se representan en color azul y rosa, respectivamente.

DISCUSIÓN

USO DE CODONES

Los resultados sugieren que el uso de codones no tiene una gran influencia sobre la eficiencia de *N*-glicosilación en la mayoría de las mutantes estudiadas. Esto puede tener varias explicaciones:

i) podría ser que los secuencias no estuvieran ubicados a la distancia justa respecto de la pausa. Si la modulación de la velocidad fuera local y en la posición donde debería ser más lenta la traducción hay residuos que no cambian apreciablemente su velocidad de decodificación, o son codificados por un codón único, no podríamos ver cambios entre una u otra variante de velocidad.

ii) podría ser que el cambio en la velocidad de traducción no tenga influencia sobre la catálisis realizada por la OST, o que posea un efecto menor y que la *N*-glicosilación, esté determinada mayoritariamente por otros factores.

iii) podría ser que el uso de codones no estuviera modulando apreciablemente la velocidad de traducción.

En el diseño de las construcciones rápida y lenta de la *cf-sfGFP* se empleó como criterio de velocidad al uso de codones, debido a que hasta ese momento sólo contábamos con referencias de literatura que mostraban una correlación entre la abundancia de codones y la concentración de ARNts en los organismos estudiados. Sin embargo, no está del todo claro aún si en humanos existe dicha correlación. En principio, de los experimentos de marcación no podemos derivar una conclusión al respecto puesto que los resultados de los diferentes experimentos no han sido concluyentes (ver más abajo). Por ello, no podemos asegurar que los cambios observados en la eficiencia de *N*-glicosilación se deban a una diferencia en la velocidad de síntesis y, solamente, podremos establecer su relación con la mera abundancia de codones. Para poder definir este punto planeamos realizar experimentos con construcciones que tengan en cuenta diferentes criterios que pudieran modular la velocidad de traducción:

- número de ARNt cognate.
- cociente entre ARNt cognate y ARNt near cognate.
- interacciones codón-anticodón según Watson-Crick vs wobble.
- autocorrelación en la aparición de codones

De los secuencias que mostraron algún cambio entre variantes, pudo observarse un aumento de la eficiencia de *N*-glicosilación para la mutante N146S con codones poco abundantes (**figura 75**), tanto en células HeLa (11 %, **tabla 3**), en HEK293 (18 %, **tabla 4**) y en COS-7 (16 %, **tabla 7**).

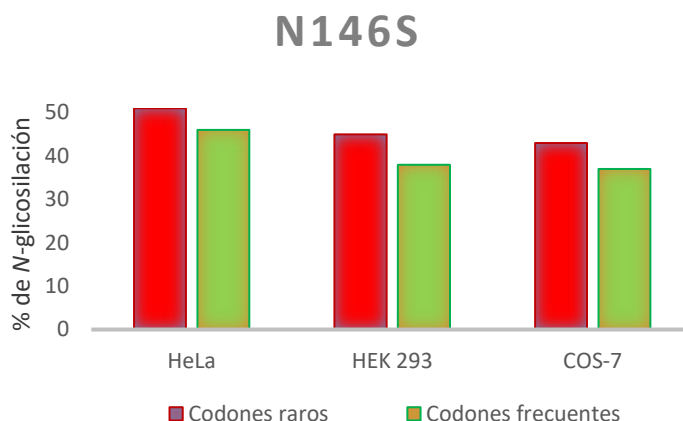


Figura 75. Porcentaje de *N*-glicosilación de la mutante N146S en las diferentes líneas celulares. En rojo se grafica la versión con codones poco frecuentes y en verde la de codones frecuentes.

En el caso de la mutante E172S se vio lo contrario a lo esperado (**figura 76**), siendo que se observa una mayor eficiencia de *N*-glicosilación en el caso de la proteína que posee codones más frecuentes (6% en células HeLa, **tabla 3** y 13% en células HEK293, **tabla 5**).

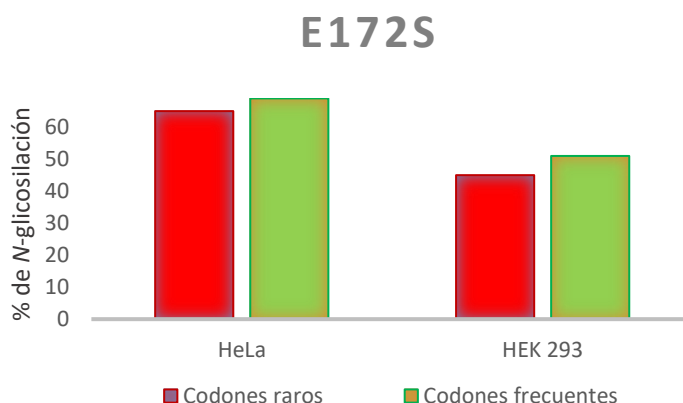


Figura 76. Porcentaje de *N*-glicosilación de la mutante E172S en las diferentes líneas celulares. En rojo se grafica la versión con codones poco frecuentes y en verde la de codones frecuentes.

Por otra parte, en células HEK293 la mutante K2 mostró una mejora en la eficiencia de *N*-glicosilación del 16%, mientras que en la mutante K1 hubo un 40% de aumento de la versión con los codones raros respecto de aquella con los codones frecuentes (**tabla 9**).

Claramente la mutante K1 merece ser estudiada en profundidad dado que, de todas las mutantes estudiadas, es la que presenta la mayor diferencia entre versiones (**figura 77**).

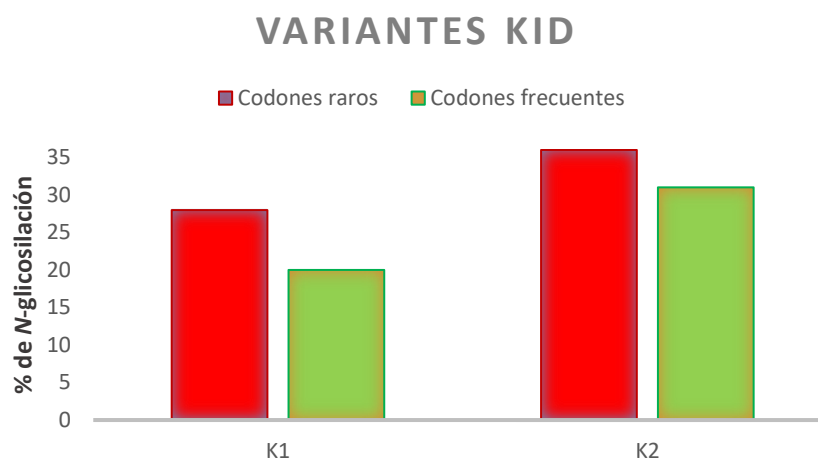


Figura 77. Porcentaje de *N*-glicosilación de las mutantes K1 y K2 en células HEK293. En rojo se grafica la versión con codones poco frecuentes y en verde la de codones frecuentes.

¿A qué podría deberse que en esta mutante se observe un cambio tan significativo? Sin lugar a dudas primero es una cuestión de identidad del secuéon, ya que mientras que K2 (cuya identidad es NGSA) sólo muestra un aumento del 16%, K1 (cuya secuencia es NGSD) que se ubica en la misma posición de la proteína aumenta en un 40%. Por otro lado, al considerar que el secuéon estará accesible al sitio activo de la oligosacariltransferasa cuando el ribosoma esté leyendo al ARNm en la secuencia correspondiente a aminoácidos ubicados a 65-75 codones de distancia hacia el extremo 3', podemos analizar la composición y coincidencia de los mismos en base a un trabajo publicado en 2014⁸¹, donde cuantifican el número de copias génicas de los ARNt en humanos. Al realizar la correspondencia entre los ARNt más frecuentes según el trabajo de Iben *et al.* y nuestra tabla de uso de codones en humanos, pudimos establecer los casos donde los codones más frecuentes se correspondían con los ARNt más abundantes y para cuáles codones raros había una correspondencia con los ARNt menos frecuentes. Casualmente, en la posición donde el ribosoma estaría ubicado, hay una coincidencia de 8/10 (**figura 78**) en la elección del codón poco abundante y el ARNt menos frecuente. Así, considerando que aquel sitio que posea menos ARNt que lo decodifiquen demorará más tiempo en ser traducido, podría explicarse el alto porcentaje de aumento en la eficiencia de *N*-glicosilación observado en la variante con codones poco frecuentes que, en este caso, también sería de lenta traducción local en el momento en que se efectuaría la modificación post traduccional.

Mutante K1

MDSKGSSQKGSRLLLLVSNNLLCQGVVSTPLESKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGD
ATNGKLTlkFISTTGKLPVPWPTLVTTLTyGVQMFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDD
GTyKTRAEVKFEGDTLVNRIEL**NGSD**FKEDGNILGHKLEYNFNshNVYITADKQKNGIKANFKIRHNV
EDGSVQLADHYQQNTPIGD**GPVLLPDNH**YLSTQSVLSKDPNEKRDHMLLEFVTAAGITHGMDE
LYKEFYPYDVPDYAGYPYDVPDYAGSYYPYDVPDYAAKDEL

Figura 78. En fucsia se muestra la ubicación del secuón de K1. Subrayada está la pausa de los últimos 63 codones y, dentro de la misma, se resaltan los residuos que estarían a una distancia de 65-75 respecto del secuón y, de estos, en verde se muestran los que muestran una coincidencia entre abundancia de codón y de ARNt.

Un análisis similar relacionando el uso de codones y la correspondencia con los ARNt para la mutante E172S (**figura 79**) muestra que de los aminoácidos que estarían a 65-75 residuos de distancia respecto de la asparagina del secuón (ELYKEFYYPYDV), ELY son residuos que en la construcción de codones poco abundantes se corresponden con ARNt poco abundantes y los codones frecuentes se correlacionan con los ARNt frecuentes en la versión rápida. Pero la secuencia EFYPYDV es común a ambas construcciones ya que forma parte del HA, siendo de lectura lenta los resaltados en rojo y de rápida los resaltados en verde (en base a que hay coincidencia entre los codones usados y los ARNt según el trabajo de Iben). Entonces, en este caso, los únicos residuos que podrían modular la velocidad de síntesis serían ELYK. Al comparar nuestra tabla de uso de codones con la de abundancia de ARNt de Iben, notamos que para la K nuestra elección de codón menos abundante se corresponde con el ARNt más frecuente. Ese residuo lisina se encuentra a 68 aminoácidos de distancia de la asparagina del secuón de la mutante E172S, lo cual nos podría estar dando un indicio de la posición donde podría estar ubicado el ribosoma en el momento de la modificación, y eso explicaría por qué es la supuesta variante rápida la que se glicosila más que la lenta (**figura 76**). Dicho de otro modo, posiblemente en este caso a nivel local lo que creemos que es la variante rápida es en realidad la lenta, y viceversa.

Mutante E172S

MDSKGSSQKGSRLLLLVSNNLLCQGVVSTPLESKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGD
ATNGKLTCLKFISTTGKLPVPWPTLVTTLTYGVMFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFSHNVYITADKQKNGIKANFKIRH**NVS**
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMLLEFVTAAGITHGMD**ELY**
KEFY**PDV**PDYAGYPYDVPDYAGSYPDVDPDYAAKDEL

Figura 79. En anaranjado se muestra la ubicación del secuón de E172S. Subrayada está la pausa de los últimos 63 codones. Se resaltan los residuos que estarían a una distancia de 65-75 respecto del secuón.

Por último, si miramos la correlación entre el uso de codones que empleamos para nuestras construcciones con la abundancia de ARNt para los residuos que se encuentran entre 65-75 codones río abajo del secuón N146S (DPNEKRDHML), vemos que para PNE existe correlación y que esa E se encuentra en la posición 68 respecto de la asparagina del secuón.

Mutante N146S

MDSKGSSQKGSRLLLLVSNNLLCQGVVSTPLESKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGD
ATNGKLTCLKFISTTGKLPVPWPTLVTTLTYGVMFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEY**NFSS****HNVYITADKQKNGIKANFKIRHNVS**
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSK**DPNEKR****DHML**LLEFVTAAGITHGMD**EL**
YKEF**YPYDVPDYAGYPYDVPDYAGSYPDVDPDYAA**KDEL

Figura 80. En violeta se muestra la ubicación del secuón de N146S. Subrayada está la pausa de los últimos 63 codones y, dentro de la misma, se resaltan los residuos que estarían a una distancia de 65-75 respecto del secuón. De estos, en verde se indican los que muestran una coincidencia entre abundancia de codón y de ARNt. En el caso de la K hemos visto que hay una correlación inversa, la M es de codón único, para H no hay datos y en el caso de la R no está claro.

Cuando analizamos qué es lo que sucede con la mutante D173N (tablas 6 y 7), vemos que el cambio es muy bajo entre variantes de uso de codones (4%), lo cual podría ser explicado por su ubicación respecto del segmento de codones poco frecuentes (figura 81), ya que el mismo finaliza en lo que sería una distancia de 65 aminoácidos respecto de la asparagina del secuón, ubicándose en la ya mencionada posición 68, uno de los aminoácidos comunes del HA.

Mutante D173N

MDSKGSSQKGSRLLLLVSNNLLCQGVVSTPLESKGEELFTGVVPILVELDGDVNNGHKFSVRGEGEGD
ATNGKLTCLKFISTTGKLPVPWPTLVTTLTGYVQMFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFSHNVYITADKQKNGIKANFKIRHNVS
NGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMLLEFVTAAGITHGMDELYK
EFYPYDVPDYAGYPYDVPDYAGSYPDVPDYAAKDEL

Figura 81. En fucsia se muestra la ubicación del secuón de D173N. Subrayada está la pausa de los últimos 63 codones y, se resaltan en amarillo a los residuos que estarían a una distancia de 65-75 respecto del secuón.

En conclusión, de esta primera parte podemos decir que de haber una modulación de la velocidad debida a una correlación entre la abundancia de codones y la cantidad de ARNt en humanos, la posición 68 respecto de la asparagina del secuón parecería estar teniendo una clara influencia en la eficiencia de *N*-glicosilación. Por otra parte, los nuevos datos de disponibilidad de ARNt permitirán diseñar construcciones que modulen de forma más efectiva la velocidad de traducción.

VELOCIDAD DE SÍNTESIS

En cuanto a los experimentos para determinar la velocidad de síntesis entre las variantes rápida y lenta, más allá de la no congruencia e inverosimilitud de los resultados, podría ser que la técnica aplicada no sea la adecuada para este modelo. El segmento en donde se cambia el uso de codones abarca sólo el ~20 % de la proteína, aún de existir un cambio de dos veces en la velocidad de lectura de este segmento, esto cambiaría el tiempo total de lectura del ARNm en cerca del 20 %. Estas diferencias son muy difíciles de detectar en experimentos de pulso y caza. Una metodología alternativa para definir este efecto podría ser medir la densidad de ocupación de ribosomas a lo largo de los mARNs. Esto se puede hacer mediante el empleo de una variante de secuenciación llamada RiboSeq, una técnica que manejan muy pocos grupos en el mundo. Otra posibilidad, sería clonar únicamente a esos 63 residuos y realizar la marcación con ³⁵S.

Decidimos realizar las marcaciones en medio sin suero debido a que, en una prueba previa, la cantidad de marca incorporada había sido mayor en la marcación total de células cuando no había suero en el medio. Pero posiblemente haya sido un error debido a que se sabe que las células responden a la falta de suero mediante una alteración de procesos metabólicos, que incluyen a la tasa de síntesis proteica y un aumento en la degradación de proteínas. Se ha visto que la tasa de elongación en células HeLa privadas de suero durante 24 hs es la mitad de

la de células cultivadas con suero⁸². En nuestros experimentos el tiempo durante el cual las células están sin suero es menor a las 3 hs, pero igual podría estar generando alguna respuesta celular que no nos permitiese ver el resultado que deberíamos si hiciéramos la marcación en presencia de suero. De todos modos, los resultados que observamos para la mutante no glicosilada “rápida” no concuerdan con una disminución de la velocidad de síntesis, sino que, contrariamente, la tasa de traducción calculada es súper rápida. Más allá de lo observado, en futuras marcaciones deberemos tener en cuenta este punto y, al menos, realizar los experimentos en ambas condiciones (con o sin suero).

HEK293 KNOCKOUT

Una observación general en esta serie de experimentos es que el grado de ocupación en las células wt siempre es mayor o igual al observado en cualquiera de las líneas KO. Algo esperable puesto que la línea wt dispone de ambas maquinarias enzimáticas.

En la mutante N146S se observó que al ser transfectada en las células B^{-/-} (**tabla 10**) se obtiene el mismo aumento (18%) en la glicosilación de la variante lenta respecto de la rápida que el obtenido anteriormente en células HEK293 wt (**tabla 4**). Esto indicaría que la subunidad STT3A podría estar supliendo la falta de B o que el sitio es modificado sólo co-traduccionalmente. Pero en el caso de las células A^{-/-} vemos que no hay cambio entre lento o rápido pero que sin embargo la subunidad STT3B puede realizar la modificación, aunque en menor grado (**figura 82**). Este resultado sugiere que la subunidad STT3B en ausencia de la STT3A se está acoplando al translocón o que permanece desacoplada del Sec61 y mientras la proteína es sintetizada modifica al secunón co-traduccionalmente, o que lo hace de modo post traduccional porque de alguna manera el sitio permanece accesible.

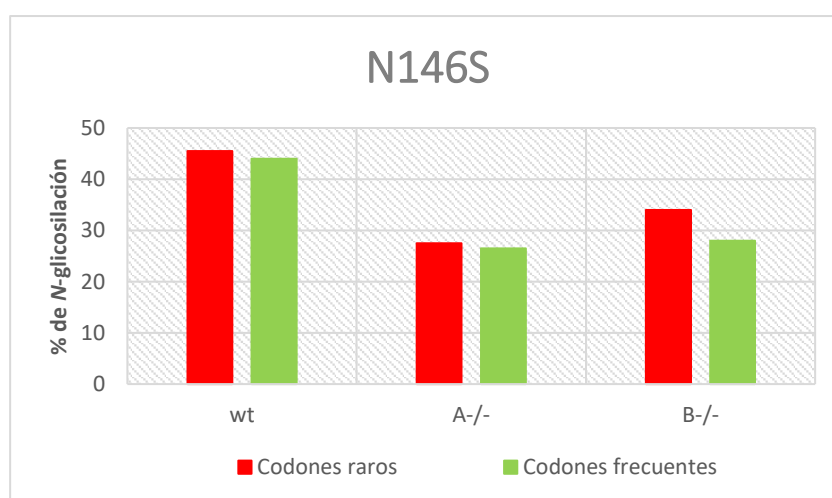


Figura 82. Porcentaje de N-glicosilación de la mutante N146S en las diferentes líneas celulares. En rojo se grafica la versión con codones poco frecuentes y en verde la de codones frecuentes.

Un trabajo previo determinó que la STT3B no se asocia con el translocón Sec61⁵², sin embargo el mismo fue realizado en células que poseen ambas subunidades catalíticas de la OST. Esto abre la posibilidad de que la STT3B pueda asociarse al translocón en ausencia de la STT3A. De hecho, luego de la resolución de la estructura de la OST de mamíferos se vió que algunos complejos de translocación no poseen una OST asociada y, se ha sugerido que esto podría ser porque la asociación se da cuando la proteína está siendo translocada dentro del lumen del RE ó porque dentro de las células existen dos poblaciones de complejos de translocación, donde solamente algunos tienen una OST asociada⁸³. Además, se ha determinado que la interacción entre la STT3A y el Sec61 tiene lugar a través de dos proteínas pequeñas de membrana (DC2 y KCP2) y que DC2 está presente también en los complejos que tienen STT3B⁸⁴. Sin embargo, en un trabajo de 2017 se establece que los complejos con STT3B ocuparían una posición distante del translocón, incluso cuando la interacción entre STT3A y el Sec61 es perturbada⁸⁵.

En el caso de la mutante NISE, solamente en las células A^{-/-} se observa la aparición de proteína no glicosilada (**figura 83**). O sea que la STT3B sola no alcanza para glicosilar al 100%, lo cual indicaría que al faltar la subunidad que realizaría la modificación de modo co-traduccional, la STT3B no puede suplir completamente su falta y comienza a aparecer proteína sin glicano (**figuras 35 y 36**).

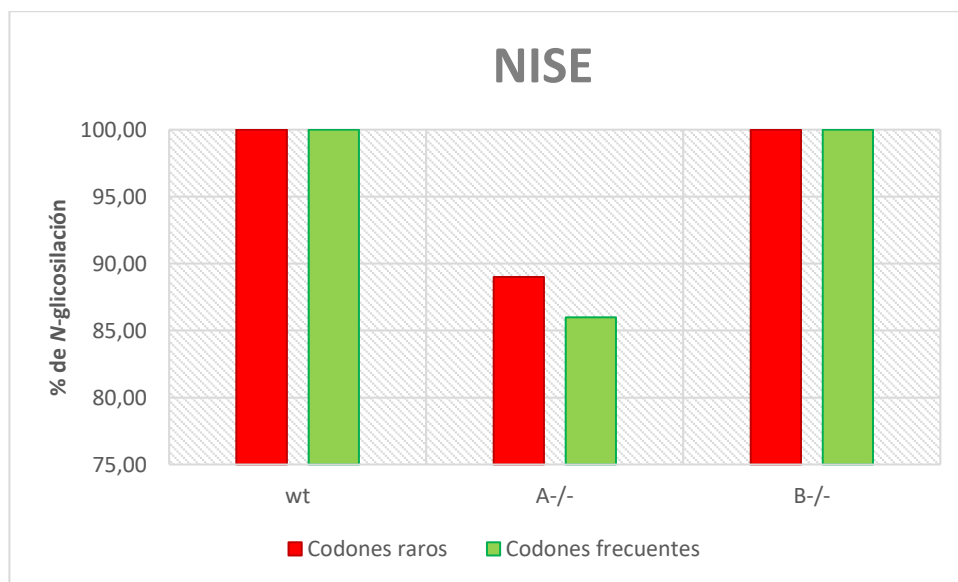


Figura 83. Porcentaje de *N*-glicosilación de la mutante NISE en las diferentes líneas celulares. En rojo se grafica la versión con codones poco frecuentes y en verde la de codones frecuentes.

Al analizar los datos de la mutante DV (**tabla 14**), vimos que cuando sólo está presente la subunidad STT3B, la glicosilación es un 20% mayor que cuando sólo está la STT3A. Esto va de la mano con las observaciones de los trabajos de Gilmore donde se enuncia que la subunidad

STT3B es más activa que la STT3A. Pero nuevamente, no sabemos si lo que está sucediendo es que la modificación está siendo realizada de modo post traduccional o si la STT3B está actuando co-traduccionalmente, ya sea, acoplándose al translocon o no.

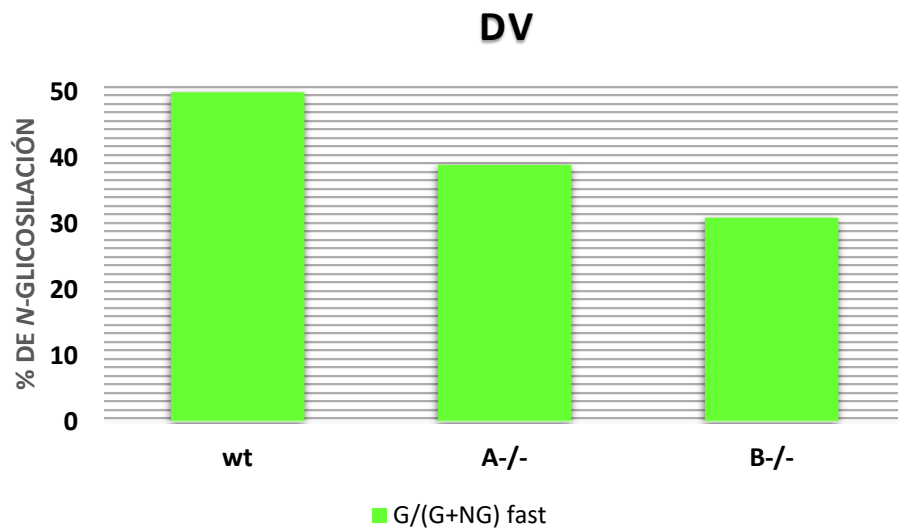


Figura 84. Porcentaje de *N*-glicosilación de la mutante DV con codones abundantes en las diferentes líneas celulares.

En el caso de la mutante E172S (**tabla 15**) cuando no está la subunidad A deja de verse el cambio en la eficiencia de *N*-glicosilación entre las variantes rápida y lenta que existe en las células HEK293 wt (**figura 85**). Por lo cual, es posible que el fenómeno esté siendo regulado por la STT3A de modo co-traduccional y que, la glicosilación observada se deba a la STT3B no acoplada al Sec61 (co o post-traduccionalmente).

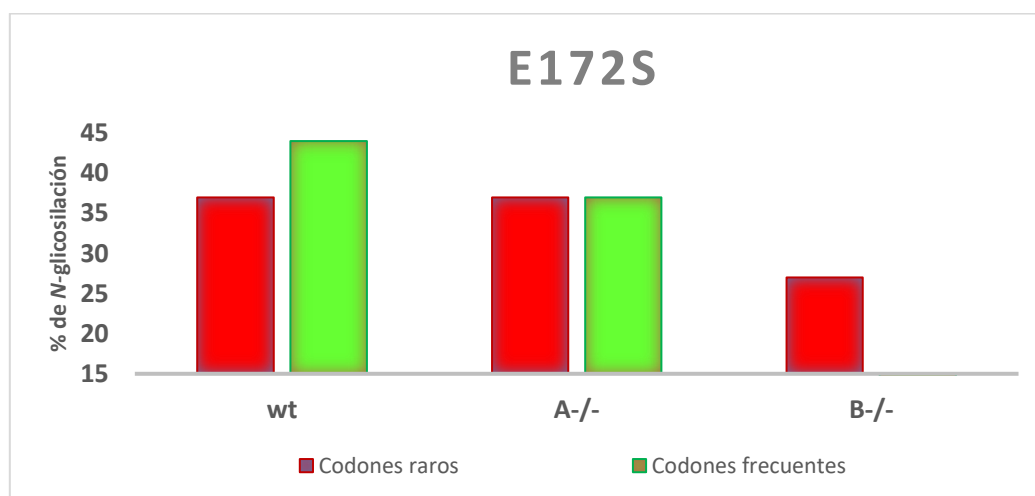


Figura 85. Porcentaje de *N*-glicosilación de la mutante E172S en las diferentes líneas celulares. En rojo se grafica la versión con codones poco frecuentes y en verde la de codones frecuentes.

En las transfecciones de la mutante N-terminal, G228N+T230S (**tabla 18**) observamos que en las células que carecen de la subunidad STT3B es donde hay un mayor porcentaje de proteína no glicosilada y que cuando no está presente la subunidad STT3A la eficiencia de glicosilación es la misma que la alcanzada en células wt (**figura 86**). Esta insensibilidad a la falta de la subunidad STT3A ya ha sido descrita en otros trabajos⁸⁶ para sitios ubicados a menos de 55 residuos de la finalización de la cadena proteica. Que no se hayan detectado cambios entre las versiones lenta y rápida, sumado a la disminución en el nivel de glicosilación en las células sin STT3B, nos habla del carácter post traduccional del secuéon.

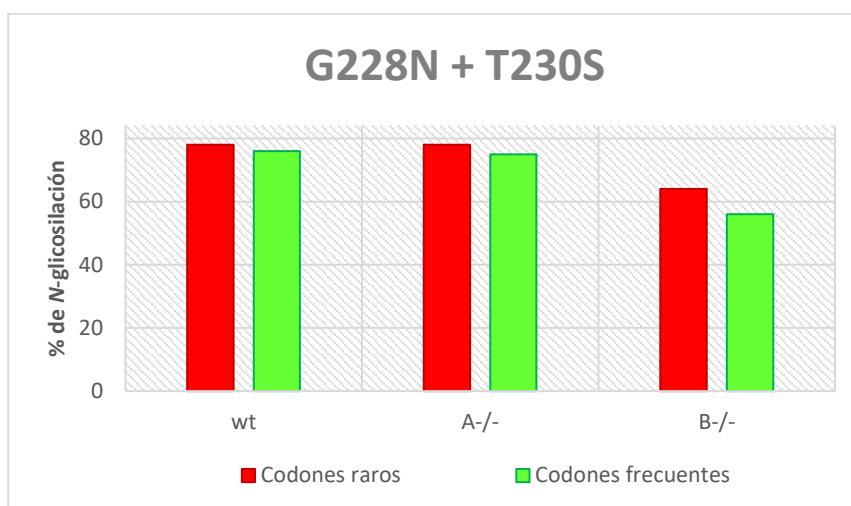


Figura 86. Porcentaje de N-glicosilación de la mutante G228N + T230S en las diferentes líneas celulares. En rojo se grafica la versión con codones poco frecuentes y en verde la de codones frecuentes.

Se ha descrito que sólo la STT3B es capaz de mediar la modificación post traduccional de los sitios ubicados a menos de 55 residuos del extremo carboxilo terminal⁵², sin embargo, en la **figura 86** se puede observar que aún cuando la subunidad STT3B no está presente, existe N-glicosilación y, sabiendo que el sitio por su ubicación en la cadena aminoacídica (a 48 residuos del final de la proteína) solamente podrá ser modificado post traduccionalmente (recordemos que los residuos que están a menos de 55 aminoácidos del C-terminal no son glicosilados antes de que finalice la síntesis de la cadena polipeptídica), estaríamos ante la presencia de una N-glicosilación post traduccional mediada por la STT3A (**figura 87B**).

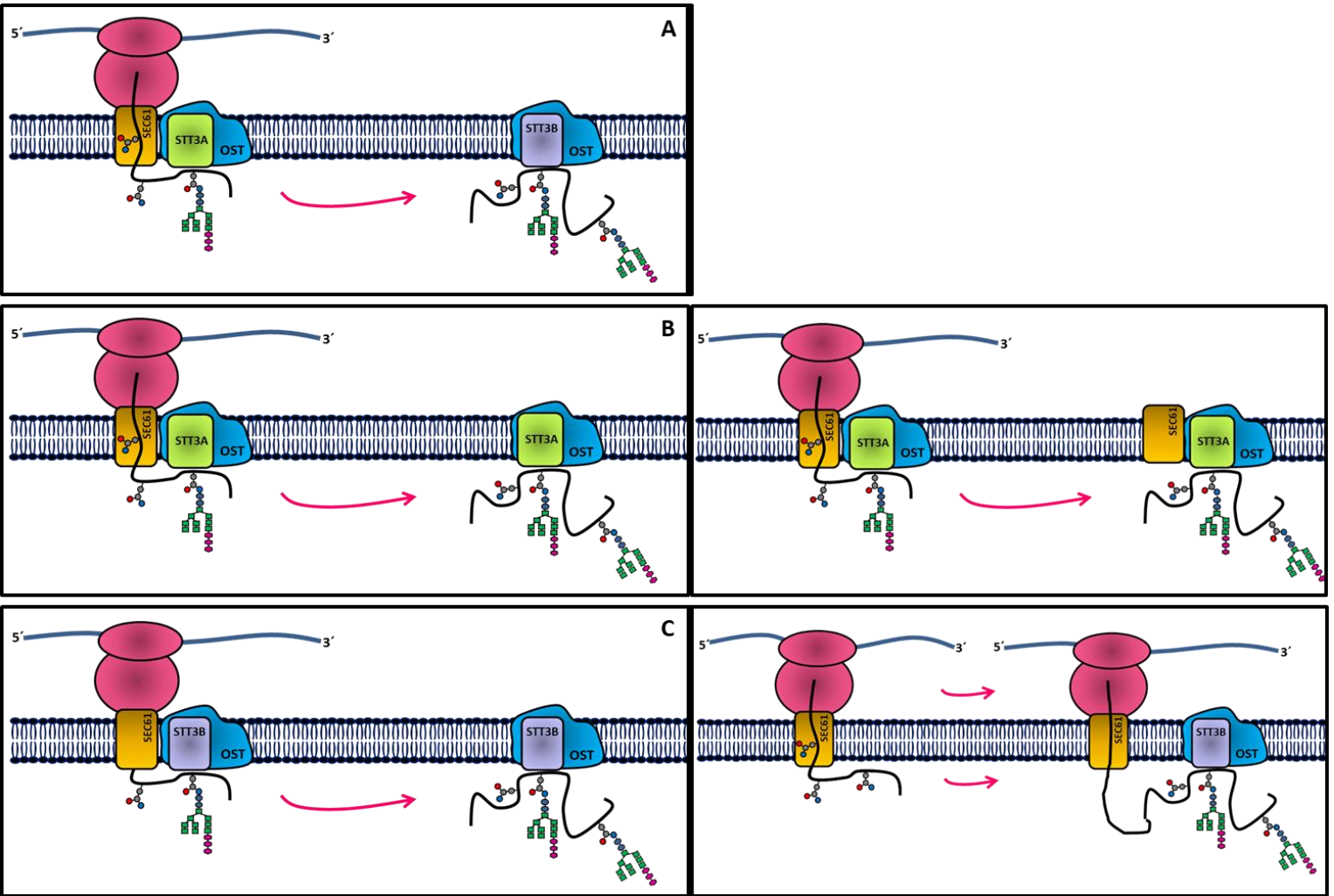


Figura 87. Se esquematiza en A la distribución de las subunidades catalíticas de la OST en células wt. En B se muestra la posible distribución de la STT3A en células sin STT3B, donde la A podría modificar post traduccionalmente a los secuones. En C se grafica la posible distribución de la STT3B en células A^{-/-}, pudiendo la misma modificar a los secuones de modo co-traduaccional.

Por último, los sitios Y151S, L7N y L7N + T8S mostraron estar 100% glicosilados en ambas versiones de uso de codones y en las tres líneas celulares HEK293 (**figuras 41 y 43**). Así, tanto la STT3A como la STT3B pueden modificar a estos sitios eficientemente.

En conclusión, si pudiéramos obtener geles cuantificables con poco ruido de fondo y buena intensidad de banda proteica en experimentos de pulso y caza, podríamos determinar en aquellos casos donde planteamos la duda, de si la subunidad STT3B está modificando co-traduccionalmente en ausencia de la STT3A y si, a la inversa, la subunidad STT3A puede modificar post traduccionalmente a los secuones en ausencia de la subunidad STT3B.

INFLUENCIA DE LA IDENTIDAD DEL RESIDUO 70.

Al restituir la cisteína 70 que había sido mutada por metionina, y comparar la eficiencia de *N*-glicosilación en las diferentes líneas celulares de HEK293 observamos que existe una mejora notable en la glicosilación cuando hay metionina en la posición 70 de la proteína en la mayoría de las mutantes. Este fenómeno resultó ser muy dependiente del secuen (identidad y posición) y del tipo celular (**figura 88**). Notable es el caso de las mutantes Y151S, NISA y NISE (**figura 49**), las cuales están casi contiguas en la estructura primaria. Mientras que en la Y151S se observa una eficiencia máxima de *N*-glicosilación tanto en presencia de C70 como de M70 en todas las líneas celulares, en las variantes M70 con los secuenes NISA y NISE se obtiene un aumento de la *N*-glicosilación del 670 % y 225 % respectivamente al ser expresadas en células *A*^{-/-}. También pudo observarse un aumento del 53% en la *N*-glicosilación de la mutante N146S con M70 en las células sin STT3A. En esta misma línea celular y en las HEK293 wt, la mutante K41S no mostró diferencias entre poseer cisteína o metionina en posición 70 (**figura 51**). Para la mutante E172S (**figura 50**) se calculó un aumento de la *N*-glicosilación del 58% en la línea HEK293 wt y, por último, la mutante C-terminal G228N + T230S (**figura 52**) se mostró más glicosilada en presencia de M70 en todas las líneas celulares HEK293.

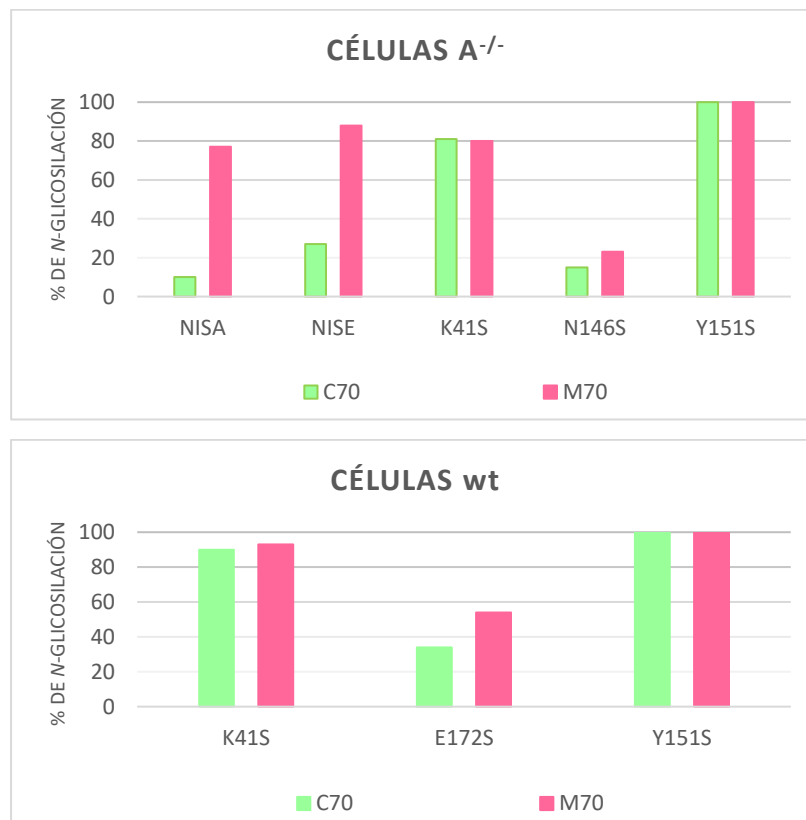


Figura 88. Porcentajes de *N*-glicosilación de las diferentes mutantes con cisteína 70 (barras verdes) ó metionina 70 (barras rosas), expresadas en HEK293 *A*^{-/-} (panel superior) ó wt (panel inferior). En verde se muestra la variante con C70 y en rosa la que tiene M70.

Se pudo descartar que la disminución en la eficiencia de *N*-glicosilación en las mutantes con cisteína se debiera a un proceso de oligomerización por formación de puentes disulfuro entre proteínas (**figura 53**). Se procedió a estudiar la estabilidad y la dinámica de plegado de las mutantes no glicosiladas con el objetivo de explicar el aumento en la eficiencia del proceso cuando hay una metionina en lugar de cisteína. Así, se pudo observar que la variante con metionina posee una menor estabilidad térmica respecto de la que tiene cisteína en posición 70 ($T_{mMet} = 83.8 \pm 0.3 \text{ }^{\circ}\text{C}$ vs $T_{mCys} = 92.0 \pm 0.4 \text{ }^{\circ}\text{C}$) y que se despliega a una menor concentración de agente desnaturante (**figura 62**). Al comparar las cinéticas de desplegado se observó que la mutante con metionina 70 se despliega con una cinética mucho más rápida que la que tiene cisteína (**figura 64**), mientras que la velocidad de replegado resultó muy parecida para ambas (**figura 66**).

En resumen, la mutación C70M produjo cambios notorios en la eficiencia de glicosilación, los cuales fueron muy dependientes de la posición de cada secunón. Los estudios de estabilidad térmica y frente a GdnHCl muestran que la mutación provoca una gran desestabilización de la proteína. A pesar de esto, el plegado global de la misma parece conservarse. Esto se evidencia por la gran similitud de los espectros de dicroísmo circular y de fluorescencia. La generación del cromóforo de las GFPs es un paso muy sensible a cambios estructurales, y el hecho de que el cromóforo madure dando un espectro idéntico en ambas proteínas evidencia que la mutación no afecta el plegado global. El estudio de la cinética de plegado y desplegado mostró que la diferencia de estabilidad se debe principalmente a una mayor velocidad de desplegado de la variante M70. Esto podría ser una consecuencia de ligeros impedimentos estéricos generados por el mayor volumen del residuo introducido en la mutación. Una disminución en la energía de interacción intramolecular del estado nativo en la mutante bajaría la energía de activación del desplegado, generando un aumento en la velocidad del proceso, con la consiguiente disminución en la estabilidad termodinámica de la proteína. Por otra parte, el replegado de estas proteínas no mostró grandes diferencias. En ambos casos este proceso mostró ser muy complejo, algo normal para proteínas con tantos residuos de prolina. En términos generales la cinética de replegado está dominada por la topología del estado nativo. Puesto que el estado nativo de ambas proteínas es idéntico, no es sorprendente que presenten cinéticas similares de plegado. De todas formas, para obtener un panorama más completo sobre este punto sería necesario un estudio biofísico de todas las variantes empleadas, y no sólo de las proteínas madre sin secunones.

De las 7 variantes de glicosilación analizadas en ambas versiones, tres vieron aumentada notablemente la ocupación de los secunones en la mutante C70M (NISA, NISE y G228N+T230S),

dos tuvieron efectos intermedios (E172S y N146S), y las otras dos mostraron poco o ningún cambio (K41S e Y151S) (**tabla 28**).

Variante	Wt	A ^{-/-}	B ^{-/-}
K41S	similar, ~100 %	similar, ~100 %	similar, ~100 %
N146S	no cuantificable	>	no cuantificable
NISA	similar, ~100 %	>>>	similar, ~100 %
NISE	similar, ~100 %	>>>	similar, ~100 %
Y151S	similar, ~100 %	similar, ~100 %	similar, ~100 %
E172S	>	no cuantificable	no cuantificable
G228N+T230S	>>	>>	>>

Tabla 28. Resumen de los cambios observados en la ocupación de los secuones al introducir la mutación C70M y ser expresadas en las líneas HEK293. “Similar, ~100 %”: variantes que presentaron niveles de ocupación cercana al 100 % y similar entre las versiones C70 y M70. Los signos > indican el grado de aumento cualitativo en la ocupación de la variante M70 respecto de la proteína C70. “no cuantificable”: la intensidad de las bandas del western blot fue distinta entre las muestras y no permite una cuantificación confiable.

Estos cambios se pueden visualizar en la **figura 89**, en donde se marca la posición de la M70 y se emplea un código de colores para clasificar los cambios en la ocupación de los secuones.

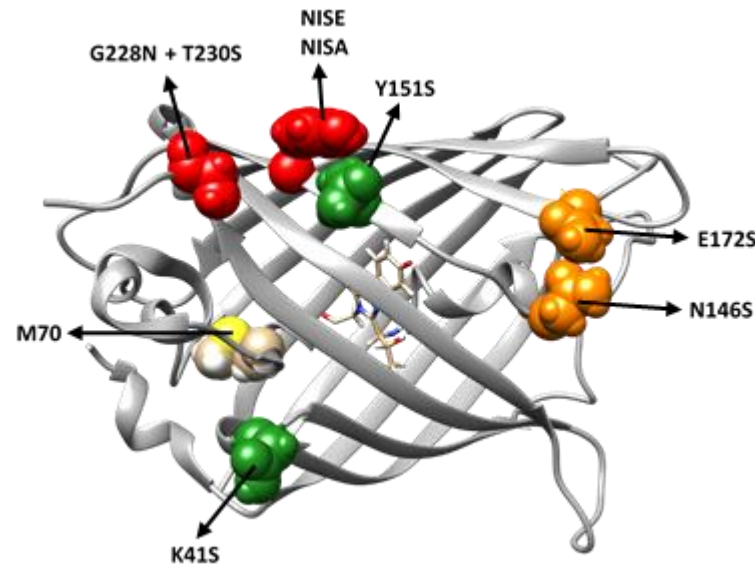


Figura 89. Ubicación de las posiciones glicosilables empleadas en las variantes C70 y M70. El código de colores indica cualitativamente los cambios en la ocupación de los secuones al introducir la mutación C70M. Rojo: gran aumento. Naranja: aumento apreciable. Verde: ocupación cercana al 100 % en todos los casos.

Las variantes K41S e Y151S presentan eficiencias de glicosilación muy altas ya en la versión C70, habiendo muy poco margen para que este parámetro mejore en la mutante C70M. Las variantes que más vieron incrementada la ocupación en la mutante C70M fueron las NISA,

NISE y G228N. Estas posiciones quedan relativamente cerca de la mutación desestabilizadora. El residuo M70 se ubica en el core hidrofóbico de la proteína, presentando una gran restricción conformational. En esa región existe un estrecho entramado de contactos, con poco espacio para acomodar cambios que impliquen el aumento de volumen de un residuo. En este escenario, la introducción de la mutación C70M puede afectar la calidad de las interacciones de los aminoácidos donde están los secuencias más afectados. Esta red de interacciones está conformada por los residuos M70, F84, I152 y H199. Estos últimos residuos están vecinos a los secuencias más afectados (**figura 90**). Por otra parte, las secuencias que mostraron una mejora en su ocupación pero en una magnitud mucho menor (N146S y E172S) se encuentran lejos de la zona perturbada.

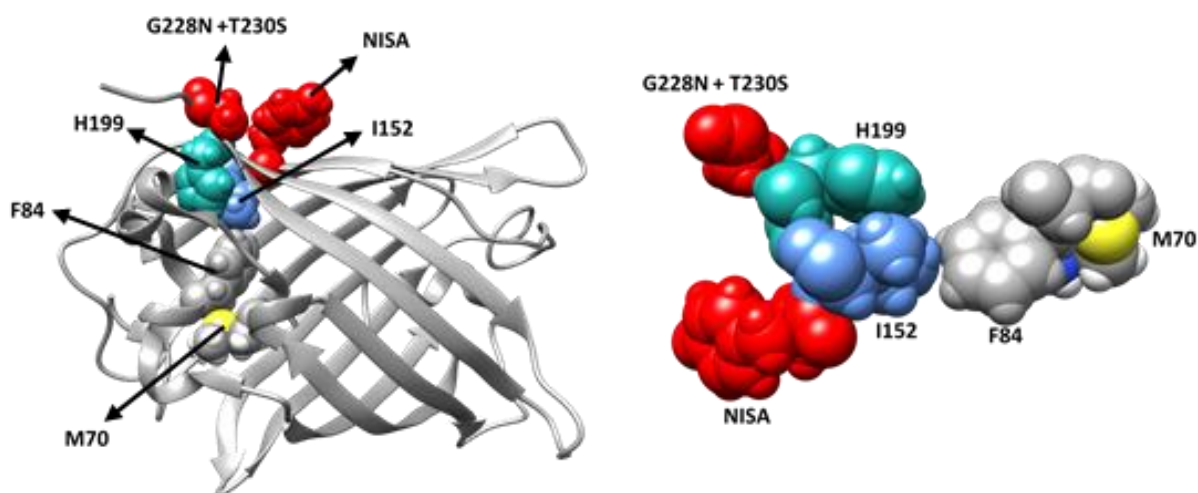


Figura 90. Contactos entre el residuo M70 y el entorno de los secuencias más afectados.

Estos resultados indican que una perturbación estructural local y relativamente suave, puesto que no altera apreciablemente la conformación del estado plegado, puede afectar drásticamente la eficiencia de ocupación de sitios de *N*-glicosilación. Para tener una visión más completa de este fenómeno sería necesario generar nuevas mutaciones que modulen la estabilidad en ambos sentidos, y estudiar si se mantiene la correlación entre perturbaciones locales y distancia a los secuencias más afectados.

Los experimentos de pulso y caza hubiesen sido de gran aporte para nuestro trabajo ya que, al haber determinado que una mutación puntual como la M70C cambia la cinética de desplegado y, en consecuencia, impacta en el estado de *N*-glicosilación, puede ser que en alguna o todas las mutantes construidas con secuencias estuviera ocurriendo un proceso de desestabilización como tal y que el patrón final de glicosilación que observamos no refleje lo que

ocurrió de modo co-traduccional sino que sea un producto de la oligosacariltransferasa actuando de modo post traducciona. Dado que proponemos un modelo (**figura 87 y tabla 29**) donde ambos tipos de células KO pueden glicosilar post traducciona, la diferencia que pudiera haber en la eficiencia de *N*-glicosilación en función de la abundancia de codones, quedaría enmascarada por el favorecimiento de una cinética de desplegamiento mayor, donde la oligosacariltransferasa podría modificar al secun post traducciona. Para evaluar esta posibilidad, debemos clonar en un vector de expresión en bacterias a cada una de las mutantes de *N*-glicosilación y realizar los ensayos de cinética mostrados en este trabajo. Otra alternativa sería tener a las mutantes de *N*-glicosilación en sus variantes de codones frecuentes/raros en el contexto de cisteína 70.

<i>N</i>-glicosilación	Co-traducciona	Post-traducciona
STT3A	✓	✓
STT3B	pobre	✓

Tabla 29. Actividades propuestas para las subunidades catalíticas de la OST.

Retomando el análisis del primer párrafo de esta parte de la discusión, vimos que la mutante G228N+T230S, cuyo secun será modificado post traducciona por estar a menos de 50 residuos del final de la proteína, muestra una mejora en la eficiencia de *N*-glicosilación cuando en su posición 70 posee metionina en lugar de cisteína y, hemos dicho que esta mejora viene dada por un aumento en la velocidad de desplegamiento, lo que le daría más tiempo a la OST para realizar la catálisis enzimática. Entonces, al evidenciar ese efecto en las células B^{-/-}, quedaría establecido que la STT3A puede modificar al sitio de modo post traducciona. En el caso de las mutantes NISA y NISE hemos visto que en ausencia de la STT3B, la subunidad STT3A puede glicosilar al 100%, pero que cuando sólo está la STT3B el patrón de glicosilación es de doble banda y la misma se debería a un proceso post traducciona mediado por la STT3B ya que al comparar a las variantes con C o M70, se observa lo ya descrito acerca de la magnificación de la glicosilación en las proteínas con M70 debido a su disminución de la estabilidad.

Si bien los estudios biofísicos realizados hasta aquí brindan información acerca del plegado de estas proteínas, los mismos deben ser interpretados cuidadosamente a la hora de extraer conclusiones válidas *in vivo*. En particular, el estado desplegado que presenta la proteína una vez que ingresa al lumen del RE puede diferir notablemente del que se obtiene luego de incubar en GdnHCl. Cuando la proteína emerge del ribosoma presenta los 10 residuos de prolina en configuración *trans*, debiendo isomerizar a uno de ellos (P89) para adquirir la estructura

nativa. Esto es una consecuencia de la actividad catalítica del ribosoma, que construye todas las uniones peptídicas en conformación *trans*. Tal escenario es muy diferente del que se tiene en los experimentos *in vitro*. Por otra parte, la gran variedad de peptidil-prolil cis-trans isomerasas disponibles en el lumen del RE constituye un factor adicional en la complejidad del proceso *in vivo*, donde la acción de estas enzimas afecta a la cinética de plegado de la P89, pudiendo ser muy distinto de lo observado *in vitro*.

A la hora de diseñar una proteína de interés biotecnológico, no podemos dejar de lado el tema de la estabilidad. En este trabajo, hemos demostrado que una mutación puntual provoca que la proteína se despliegue más rápido, otorgándole a la OST más tiempo para realizar la transferencia del *N*-glicano. Dado que la glicosilación disminuye la movilidad estructural de la proteína e incrementa su estabilidad termodinámica⁸⁷, una vez glicosilada, la estabilidad termodinámica de la proteína se vería aumentada⁸⁸. Entonces, el estudio y conocimiento de la dinámica del plegado glicoproteico podría ser aplicado en la ingeniería de proteínas terapéuticas. Así, por ejemplo, podría generarse una mutación en la secuencia del interferón beta tal que su estabilidad termodinámica sea menor, su velocidad de desplegamiento mayor y esperar así que la proteína resulte más glicosilada, potenciando sus propiedades enzimáticas como se comentó en la introducción.

ANEXO

TABLA DE USO DE CODONES EN HUMANOS					
Aminoácidos	Codones	Frecuencia	Total Transfers Cogantes	Total Transfers Near Cognates	ratio
A	GCG	0,737	14	65	0,215385
	GCC	2,773	29	48	0,604167
	GCT	1,845	38	16	2,375
	GCA	1,582	38	32	1,1875
C	TGC	1,262	30	38	0,789474
	TGT	1,058	30	24	1,25
D	GAT	2,178	19	66	0,287879
	GAC	2,51	19	41	0,463415
E	GAA	2,896	13	55	0,236364
	GAG	3,959	26	47	0,553191
F	TTT	1,757	12	26	0,461538
	TTC	2,028	12	58	0,206897
G	GGC	2,222	15	35	0,428571
	GGG	1,647	16	49	0,326531
	GGA	1,647	9	49	0,183673
	GGT	1,075	24	47	0,510638
H	CAT	1,086	11	61	0,180328
	CAC	1,509	11	32	0,34375
I	ATA	0,749	19	57	0,333333
	ATT	1,6	27	31	0,870968
	ATC	2,082	22	65	0,338462
K	AAA	2,444	17	67	0,253731
	AAG	3,186	34	64	0,53125
L	TTG	1,293	14	26	0,538462
	CTA	0,715	15	34	0,441176
	CTT	1,319	15	27	0,555556
	CTG	3,964	13	42	0,309524
	TTA	0,767	7	29	0,241379
	CTC	1,959	12	24	0,5
M	ATG	2,204	20	55	0,363636
N	AAC	1,91	33	50	0,66
	AAT	1,696	33	58	0,568966
P	CCA	1,692	17	24	0,708333
	CCC	1,979	10	22	0,454545
	CCT	1,754	17	23	0,73913
	CCG	0,692	11	46	0,23913
Q	CAG	3,423	32	30	1,066667
	CAA	1,234	11	48	0,229167
R	AGA	1,217	6	41	0,146341
	CGG	1,142	11	42	0,261905
	CGA	0,617	13	26	0,5
	AGG	1,196	11	51	0,215686
	CGT	0,454	13	27	0,481481
	CGC	1,042	7	22	0,318182

TABLA DE USO DE CODONES EN HUMANOS (continuación)					
Aminoácidos	Codones	Frecuencia	Total Transfers Cogantes	Total Transfers Near Cognates	ratio
S	TCC	1,768	11	65	0,169231
	AGC	1,946	8	51	0,156863
	TCA	1,221	16	16	1
	TCG	0,441	9	28	0,321429
	TCT	1,522	16	5	3,2
T	AGT	1,213	8	36	0,222222
	ACA	1,511	16	34	0,470588
	ACC	1,889	10	60	0,166667
	ACT	1,312	16	21	0,761905
	ACG	0,605	12	52	0,230769
V	GTT	1,103	16	45	0,355556
	GTC	1,446	11	55	0,2
	GTG	2,812	21	36	0,583333
	GTA	0,708	16	47	0,340426
W	TGG	1,317	9	45	0,2
Y	TAC	1,531	15	45	0,333333
	TAT	1,219	15	14	1,071429

RESUMEN DE LA TABLA DE USO DE CODONES					
Aminoácido	# codones	¿Modulable?	Codón rápido 1 ^{ra} alternativa	Codón lento	Codón rápido 2 ^{da} alternativa
A	4	SI	GCC	GCG	GCT
C	2	+ o -	TGC	TGT	
D	2	+ o -	GAC	GAT	
E	2	SI	GAG	GAA	
F	2	+ o -	TTC	TTT	
G	4	SI	GGC	GGT	GGA
H	2	SI	CAC	CAT	
I	3	SI	ATC	ATA	ATT
K	2	SI	AAG	AAA	
L	6	SI	CTG	CTA	
M	1	NO	ATG	ATG	
N	2	+ o -	AAC	AAT	
P	4	SI	CCC	CCG	CCT, luego CCA
Q	2	SI	CAG	CAA	
R	6	SI	AGA	CGT	AGG
S	6	SI	AGC	TCG	TCT
T	4	SI	ACC	ACG	ACA
V	4	SI	GTG	GTA	
W	1	NO	TGG	TGG	
Y	2	+ o -	TAC	TAT	

- "Modulable" significa si existen opciones para poner rápidos o lentos.
- Donde dice "+ o -" significa que casi no hay gran diferencia en el uso de codones entre el más abundante y el menos abundante.
- La 2da alternativa para el codón rápido puede usarse para modular el contenido de GC hacia arriba o hacia abajo de los mensajeros rápidos.

Tabla generada a partir del trabajo de Iben *et al.* 2014:

Aminoácido	Codón rápido	Codón lento
A	GCT	GCG
C	nd	nd
D	nd	nd
E	GAG	GAA
F	nd	nd
G	GGC	GGG
H	nd	nd
I	ATT	ATC
K	AAA	AAG
L	CTG	CTA
M	-	-
N	AAC	AAT
P	CCT	CCG
Q	CAG	CAA
R	CGT	CGG
S	TCT	TCG
T	ACT	ACG
V	GTG	GTA
W	-	-
Y	TAC	TAT

Secuencias nucleotídicas finales de las *cf-sfGFP* no glicosiladas.

Fast *cf-sfGFP* -ER-NG (no glicosilada) *KpnI*-ATG-SP prolactin-*XhoI*-*cf-sfGFP* -*EcoRI*-3xHA-KDEL-*XbaI*

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTTCAAC
TCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCCC
TGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCTAACGAGAAGAG
AGACCACATGGTGCTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTACA
AG**GAATTC**TACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC
TCCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

Fast *cf-sfGFP* -Cit-NG (no glicosilada) *KpnI*- *cf-sfGFP* -*EcoRI*-3xHA-KDEL-*XbaI*

GGTACCATGAGCAAAGGTGAAGAACTATTTACGGGTGTAGTACCGATACTAGTAGAACTAGATGGTGA
TGTAATGGTCATAAATTTTCGGTACGTGGCGAGGGCGAGGGCGACGCTACAAACGGCAAGCTGACA
CTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGCCATGGCCAACACTGGTGACAACACTGACATAC
GGCGTGAGATGTTCTCTAGATACCCAGACCACATGAAGCAGCAGCACTTCTTCAAGTCTGCTATGCCA
GAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAGGACGACGGCACATACAAGACAAGAGCTGAGG
TGAAGTTCGAGGGCGACACACTGGTGAACAGAATTGAGCTGAAGGGCATTGACTTCAAGGAGGACGG
CAACATTCTGGGCCACAAGCTGGAGTACAACCTTCAACTCTCACAACGTGTACATTACAGCTGACAAGCA
GAAGAACGGAATTAAGGCTAACTTCAAGATTAGACACAACGTGGAGGACGGATCCGTGCAGCTGGCT
GACCACTACCAGCAGAACACACCTATCGGCGACGGCCCTGTGCTGCTGCCTGACAACCACTACCTGTCT
ACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGAGAGACCACATGGTGCTGCTGGAGTTCGTGAC
CGCTGCTGGCATCACACACGGCATGGACGAGCTGTACAAG**GAATTC**TACCCATACGATGTTCTGACT
ATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC**TCCTATCCATATGACGTTCCAGATTACGCT**
GCTAAGGACGAGCTGTAGTCTAGA

Fast *cf-sfGFP* -Secreted-NG (no glicosilada) *KpnI*-ATG-SP prolactin-*XhoI*- *cf-sfGFP* -*EcoRI*-3xHA-*XbaI*

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTTCAAC
TCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCCC
TGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCTAACGAGAAGAG
AGACCACATGGTGCTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTACA

AGGAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCGGACTATGCAGGC
TCCTATCCATATGACGTTCCAGATTACGCTGCTAGTCTAGA

Slow cf-sfGFP -ER-NG (no glicosilada) KpnI-ATG-SP prolactin-XhoI-cf-sfGFP-EcoRI-3xHA-KDEL-XbaI

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTACAACACTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCC
GGTACTACTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACG
TGATCATATGGTACTACTAGAATTTGTAACGGCGGGCGGGTATAACGCATGGTATGGATGAACTATATAA
AGGAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCGGACTATGCAGGCTC
CTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

Slow cf-sfGFP -Cit-NG (no glicosilada) KpnI- cf-sfGFP -EcoRI-3xHA-KDEL-XbaI
GGTACCATGAGCAAAGGTGAAGAACTATTTACGGGTGTAGTACCGATACTAGTAGAACTAGATGGTGA
TGTAATGGTCATAAATTTTCGGTACGTGGCGAGGGCGAGGGCGACGCTACAAACGGCAAGCTGACA
CTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGCCATGGCCAACACTGGTGACAACACTGACATAC
GGCGTGAGATGTTCTCTAGATACCCAGACCACATGAAGCAGCAGCACTTCTTCAAGTCTGCTATGCCA
GAGGGCTACGTGCAGGAGAGAACAATTTCTTTCAAGGACGACGGCACATACAAGACAAGAGCTGAGG
TGAAGTTCGAGGGCGACACACTGGTGAACAGAATTGAGCTGAAGGGCATTGACTTCAAGGAGGACGG
CAACATTCTGGGCCACAAGCTGGAGTACAACCTCAACTCTCACAACGTGTACATTACAGCTGACAAGCA
GAAGAACGGAATTAAGGCTAACTTCAAGATTAGACACAACGTGGAGGACGGATCCGTACAACACTAGCG
GATCATTATCAACAAAATACGCCGATAGGTGATGGTCCGGTACTACTACCGGATAATCATTATCTATCG
ACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACGTGATCATATGGTACTACTAGAATTTGTAACG
GCGGCGGGTATAACGCATGGTATGGATGAACTATATAAAGGAATTCACCCATACGATGTTCTGACTA
TGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGCTCTATCCATATGACGTTCCAGATTACGCTGC
TAAGGACGAGCTGTAGTCTAGA

Slow cf-sfGFP-Secreted-NG (no glicosilada) KpnI-ATG-SP prolactin-XhoI- cf-sfGFP -
EcoRI-3xHA-XbaI

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTACAACACTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCC
GGTACTACTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACG
TGATCATATGGTACTACTAGAATTTGTAACGGCGGGCGGGTATAACGCATGGTATGGATGAACTATATAA

AGAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGCTC
CTATCCATATGACGTTCCAGATTACGCTGCTAGTCTAGA

SECUENCIAS NUCLEOTÍDICAS DE LAS MUTANTES DE N-GLICOSILACIÓN

KpnI-ATG-SP prolactin-XhoI-cf-sfGFP-EcoRI-3xHA-KDEL-XbaI

(L7N: CTA7AAT) slow

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAAATTTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGCAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTACAACACTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCC
GGTACTACTACCGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACG
TGATCATATGGTACTACTAGAATTTGTAACGGCGGGCGGGTATAACGCATGGTATGGATGAACATATAA
AGAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGCTC
CTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(L7N: CTA7AAT) fast

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAAATTTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGCAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCCC
TGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGAG
AGACCACATGGTGCTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTACA
AGAGAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC
TCCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(L7N: CTA7AAT) + (T8S: ACG8TCG) slow

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAAACTTTTCGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGCAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTACAACACTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCC

GGTACTACTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACG
TGATCATATGGTACTACTAGAAATTTGTAACGGCGGGCGGGTATAACGCATGGTATGGATGAACTATATAA
AGAAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGCTC
CTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(L7N: CTA7AAT) + (T8S: ACG8TCG) fast

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAAAATTTTCGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCCC
TGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGAG
AGACCACATGGTGCTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTACA
AGAAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC
TCCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(K41S: AAG41TCC) slow

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCTCCCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTACAACCTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCC
GGTACTACTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACG
TGATCATATGGTACTACTAGAAATTTGTAACGGCGGGCGGGTATAACGCATGGTATGGATGAACTATATAA
AGAAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGCTC
CTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(K41S: AAG41TCC) fast

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCTCCCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCCC
TGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGAG
AGACCACATGGTGCTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTACA
AGAAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC
TCCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(N146S: AAC146TCC) slow

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGCAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCTCC
TCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTACAACACTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCC
GGTACTACTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACG
TGATCATATGGTACTACTAGAAATTTGTAACGGCGGGCGGGTATAACGCATGGTATGGATGAACTATATAA
GAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGCTC
CTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(N146S: AAC146TCC) fast

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGCAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTTCTCC
TCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCCC
TGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGAG
AGACCACATGGTGTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTACA
AGGAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC
TCCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(Y151S: TAC151TCC) slow

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGCAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTTCAAC
TCTCACAACGTGTCCATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTACAACACTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCC
GGTACTACTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACG
TGATCATATGGTACTACTAGAAATTTGTAACGGCGGGCGGGTATAACGCATGGTATGGATGAACTATATAA
GAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGCTC
CTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(Y151S: TAC151TCC) fast

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG

CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTCACAACGTG**TCC**ATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCCC
TGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGAG
AGACCACATGGTGCTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTACA
AG**GAATTC****TACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC**
TCCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(Y151N: TAC151AAC) slow

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTCACAACGTG**AAC**ATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACA
CAACGTGGAGGACGGATCCGTACAACCTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTC
CGGTACTACTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAAC
GTGATCATATGGTACTACTAGAATTTGTAACGGCGGCGGGTATAACGCATGGTATGGATGAAGTATAT
AAA**GAATTC****TACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGG**
CTCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(Y151N: TAC151AAC) fast

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTCACAACGTG**AAC**ATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACA
CAACGTGGAGGACGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCC
CTGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGA
GAGACCACATGGTGCTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTAC
AAG**GAATTC****TACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGG**
CTCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(Y151N: TAC151AAC + T153S: ACA153TCC) slow

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG

AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTTCAAC
TCTCACAACGTGAACATTCCGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTACAACCTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCC
GGTACTACTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACG
TGATCATATGGTACTACTAGAATTTGTAACGGCGGGCGGGTATAACGCATGGTATGGATGAACTATATAA
GAATTC**TACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGCTC**
CTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(Y151N: TAC151AAG + T153S: ACA153TCC) fast

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTTCAAC
TCTCACAACGTG**AAGATTCC**GCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACA
CAACGTGGAGGACGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCC
CTGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGA
GAGACCACATGGTGCTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTAC
AAG**GAATTC****TACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGG**
CTCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(E172S: GAG172TCC) slow

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTTCAAC
TCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGTCCGACGGATCCGTACAACCTAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCC
GGTACTACTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACG
TGATCATATGGTACTACTAGAATTTGTAACGGCGGGCGGGTATAACGCATGGTATGGATGAACTATATAA
GAATTC**TACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGCTC**
CTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(E172S: GAG172TCC) fast

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCAGCACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTTCAAC
TCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTG**TCC**GACGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCCC
TGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGAG

AGACCACATGGTGCTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTACA
AGGAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC
TCCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(D173N: GAC519AAC) slow

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGAACGGATCCGTACAACAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCC
GGTACTACTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACG
TGATCATATGGTACTACTAGAATTTGTAACGGCGGGCGGGTATAACGCATGGTATGGATGAACTATATAA
AGGAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC
CTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(D173N: GAC519AAC) fast

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGAACCGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCCC
TGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGAG
AGACCACATGGTGCTGCTGGAGTTCGTGACCGCTGCTGGCATCACACACGGCATGGACGAGCTGTACA
AGGAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC
TCCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(G228N: GGT228AAT) + (T230S: ACG230TCG) slow

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTCAAC
TCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTACAACAGCGGATCATTATCAACAAAATACGCCGATAGGTGATGGTCC
GGTACTACTACCGGATAATCATTATCTATCGACGCAATCGGTACTATCGAAAGATCCGAATGAAAAACG
TGATCATATGGTACTACTAGAATTTGTAACGGCGGGCGAATATTCGATGGTATGGATGAACTATATAA
AGGAATTCACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC
CTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

(G228N: GGT228AAC) + (T230S: ACA230TCC) fast

GGTACCATGGACAGCAAAGGTTCTGTCGAGAAAGGGTCCCGCCTGCTCCTGCTGCTGGTGGTGTCAA
ATCTACTCTTGTGCCAGGGTGTGGTCTCCACCCCCCTCGAGAGCAAAGGTGAAGAACTATTTACGGGT
GTAGTACCGATACTAGTAGAACTAGATGGTGATGTAAATGGTCATAAATTTTCGGTACGTGGCGAGGG
CGAGGGCGACGCTACAAACGGCAAGCTGACACTGAAGTTCATTTCTACAACAGGCAAGCTGCCAGTGC
CATGGCCAACACTGGTGACAACACTGACATACGGCGTGCAGATGTTCTCTAGATACCCAGACCACATGA
AGCAGCACGACTTCTTCAAGTCTGCTATGCCAGAGGGCTACGTGCAGGAGAGAGAACAATTTCTTTCAAG
GACGACGGCACATACAAGACAAGAGCTGAGGTGAAGTTCGAGGGCGACACACTGGTGAACAGAATTG
AGCTGAAGGGCATTGACTTCAAGGAGGACGGCAACATTCTGGGCCACAAGCTGGAGTACAACCTTCAAC
TCTCACAACGTGTACATTACAGCTGACAAGCAGAAGAACGGAATTAAGGCTAACTTCAAGATTAGACAC
AACGTGGAGGACGGATCCGTGCAGCTGGCTGACCACTACCAGCAGAACACACCTATCGGCGACGGCCC
TGTGCTGCTGCCTGACAACCACTACCTGTCTACCCAGTCTGTGCTGTCTAAGGACCCTAACGAGAAGAG
AGACCACATGGTGCTGCTGGAGTTCGTGACCGCTGCT**AACATCTCC**CACGGCATGGACGAGCTGTACA
AG**GAATTCTACCCATACGATGTTCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGC**
TCCTATCCATATGACGTTCCAGATTACGCTGCTAAGGACGAGCTGTAGTCTAGA

BIBLIOGRAFÍA

- (1) Scheper, G.C.; van der Knaap, M.S.; Proud, C.G. Translation matters: Protein synthesis defects in inherited disease. *Nat. Rev. Genet.* 2007, 8, 711–723
- (2) Gingold, H. and Pilpel, Y. (2011), Determinants of translation efficiency and accuracy. *Molecular Systems Biology*, 7: n/a
- (3) Grantham R, Gautier C, Gouy M, Mercier R, Pave A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* 1980;8(1):r49–62
- (4) Pop C, et al. (2014) Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol* 10:770
- (5) Crick FH. Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol.* 1966; 19:548–55
- (6) Yu, C.-H. et al. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol. Cell* 59, 744–754 (2015)
- (7) Hershberg R & Petrov DA (2008) Selection on codon bias. *Annual review of genetics* 42:287-299.
- (8) Deane CM, Saunders R. The imprint of codons on protein structure. *Biotechnol J.* 2011;6:641–9
- (9) Spencer PS, Barral JM. Genetic code redundancy and its influence on the encoded polypeptides. *Computational and Structural Biotechnology Journal.* 2012;1:e201204006
- (10) Rodnina, M. V et al., 2005. Recognition and selection of tRNA in translation. *FEBS Letters*, 579(4 SPEC. ISS.), pp.938–942
- (11) Gardin, J. et al., 2014. Measurement of average decoding rates of the 61 sense codons in vivo. *eLife*, 3, pp.1–20
- (12) Varenne S, Buc J, Llobes R, Lazdunski C (1984) Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol* 180: 549–576
- (13) Gingold, H. et al. A dual program for translation regulation in cellular proliferation and differentiation. *Cell* 158, 1281–1292 (2014)
- (14) Akashi H. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics.* 1994;136:927–935
- (15) Ikemura, T., Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes, *J. Mol. Biol.*, Volume146, Issue 1, 15 February 1981, Pages 1-21

- (16) Ikemura, T., Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system, *J. Mol. Biol.*, 151,389, 1981
- (17) dos Reis, M., Savva, R. & Wernisch, L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* 32, 5036–5044 (2004)
- (18) Man, O., Pilpel, Y.: Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. *Nat. Genet.* 39(3), 415–421 (2007)
- (19) Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. 2001. Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J. Mol. Evol.* 53:290–98
- (20) Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, Sachs MS, Liu Y. 2013. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* 495: 111–115
- (21) Duret, L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.* 16, 287–289 (2000)
- (22) Moriyama, E. N. & Powell, J. R. Codon usage bias and tRNA abundance in *Drosophila*. *J. Mol. Evol.* 45, 514–523 (1997)
- (23) Yamao F, Andachi Y, Muto A, Ikemura T, Osawa S. Levels of tRNAs in bacterial cells as affected by amino acid usage in proteins. *Nucleic Acids Res.* 1991 Nov 25;19(22):6119-22.
- (24) Camiolo, S., Farina, L. & Porceddu, A. The relation of codon bias to tissue-specific gene expression in *Arabidopsis thaliana*. *Genetics* 192, 641–649 (2012).
- (25) Nakamura M, Sugiura M. Translation efficiencies of synonymous codons are not always correlated with codon usage in tobacco chloroplasts. *Plant J.* 2007
- (26) Urrutia A.O. and Hurst, L.D. (2001) Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. *Genetics*, 159, 1191–1199
- (27) Doherty A, McInerney JO. Translational selection frequently overcomes genetic drift in shaping synonymous codon usage patterns in vertebrates. *Mol Biol Evol.* 2013;30:2263–2267.
- (28) O'Brien EP, Ciryam P, Vendruscolo M, Dobson CM (2014) Understanding the influence of codon translation rates on cotranslational protein folding. *Acc Chem Res* 47:1536–1544
- (29) Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol.* 1985;2:13–34
- (30) Koutmou, K. S., Radhakrishnan, A. & Green, R. Synthesis at the speed of codons. *Trends Biochem. Sci.* 40, 717–718 (2015).

- (31) Chu, D., Barnes, D. J. & von der Haar, T. The role of tRNA and ribosome competition in coupling the expression of different mRNAs in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 39, 6705–6714 (2011).
- (32) Rojiani MV, Jakubowski H and Goldman E (1990) Relationship between protein synthesis and concentrations of charged and uncharged tRNA^{Trp} in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the USA* 87: 1511–1515.
- (33) Pedersen S (1984) *Escherichia coli* ribosomes translate in vivo with variable rate. *EMBO J* 3: 2895–2898
- (34) Paige S, Spencer et. al. “Silent Substitutions Predictably Alter Translation Elongation Rates and Protein Folding Efficiencies”. *J. Mol. Biol.* (2012) 422, 328–335
- (35) Sørensen, M. A. & Pedersen, S. (1991). Absolute in vivo translation rates of individual codons in *Escherichia coli*: the two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J. Mol. Biol.* 222, 265–280
- (36) Dana A, Tuller T. 2014 The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* 42, 9171–9181
- (37) Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJ, Jackson SE, Wills MR, Weissman JS. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep* 8: 1365-1379, 2014.
- (38) Stadler, M. & Fire, A. (2011). Wobble base-pairing slows in vivo translation elongation in metazoans. *RNA*, 17, 2063–2073.
- (39) Clarke Tft, Clark PL. Rare codons cluster. *PLoS ONE* (2008) 3:e3412.
- (40) J.L. Chaney, P.L. Clark. Roles for synonymous codon usage in protein biogenesis. *Annu. Rev. Biophys.*, 44 (2015), pp. 143-166
- (41) Qu X, Wen JD, Lancaster L, Noller HF, Bustamante C, Tinoco I Jr. The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature*. 2011 Jul 6;475(7354):118-21.
- (42) Fredrick K, Ibba M. How the sequence of a gene can tune its translation. *Cell*. 2010 Apr 16;141(2):227-9.
- (43) Cannarozzi G, Schraudolph NN, Faty M, von Rohr P, Friberg MT, Roth AC, Gonnet P, Gonnet G, Barral Y. A role for codon order in translation dynamics. *Cell*. 2010 Apr 16;141(2):355-67.
- (44) Z.M. Petrushenko, T.V. Budkevich, V.F. Shalak, B.S. Negrutskii, A.V. El'skaya. Novel Complexes of Mammalian Translation Elongation Factor eEF1A-GDP with Uncharged ARNt and Aminoacyl-ARNt Synthetase. *Eur. J. Biochem.*, 269 (2002), pp. 4811–4818

- (45) Faure G, Ogurtsov AY, Shabalina SA, Koonin EV. Role of mRNA structure in the control of protein folding. *Nucleic Acids Res* 2016; 44:10898-911
- (46) GRISWOLD KE, MAHMOOD NA, IV, ERSON BL, GEORGIU G. Effects of codon usage versus putative 5'-mRNA structure on the expression of *Fusarium solani* cutinase in the *Escherichia coli* cytoplasm. *Protein Expr Purif*. 2003;27:134–42
- (47) G Blobel, B Dobberstein. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *The Journal of Cell Biology* Dec 1975, 67 (3) 835-851
- (48) Goochee CF, Gramer MJ, Andersen DC, Bahr JB, Rasmussen JR: The Oligosaccharides of Glycoproteins: Factors Affecting their Synthesis and their Influence on Glycoprotein Properties. In *Frontiers in Bioprocessing II*. Edited by Todd P, Sikdar SK, Bier M. Washington: American Chemical Society; 1992:199-240
- (49) Kukuruzinska MA, Lennon K. Protein N-glycosylation: molecular genetics and functional significance. *Crit Rev Oral Biol Med*. 1998;9(4):415-48.
- (50) Mohorko E, Glockshuber R, Aebersold M. Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *J Inher Metab Dis*. 2011 Aug;34(4):869-78. doi: 10.1007/s10545-011-9337-1. Epub 2011 May 26.
- (51) Kelleher DJ, Karaoglu D, Mandon EC, Gilmore R. Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. *Mol Cell*. 2003 Jul;12(1):101-11.
- (52) Ruiz-Canada, C., Kelleher, D. J. & Gilmore, R. Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell* 136, 272–283 (2009)
- (53) Shrima, S., Trueman, S. F. and Gilmore, R. (2013b). Extreme C-terminal sites are posttranslocationally glycosylated by the STT3B isoform of the OST. *J. Cell Biol.* 201, 81–95.
- (54) Benjamin Luke Schulz (2012). *Beyond the Sequon: Sites of N-Glycosylation, Glycosylation*, Dr. Stefana Petrescu (Ed.), ISBN: 978-953-51-0771-2, InTech, DOI: 10.5772/50260. Available from: <http://www.intechopen.com/books/glycosylation/beyond-the-sequon-sites-of-n-glycosylation>
- (55) Kornfeld R, Kornfeld S: Assembly of Asparagine-Linked Oligosaccharides. *Annu Rev Biochem* 1985, 54:631-664
- (56) Kasturi, L., H. Chen & S. H. Shakin-Eshleman, (1997) Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. *Biochem J* 323: 415-419.

- (57) Kasturi L, Eshleman JR, Wunner WH, Shakin-Eshleman SH. The hydroxy amino acid in an Asn-X-Ser/Thr sequon can influence N-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein. *J Biol Chem*. 1995 Jun 16;270(24):14756-61.
- (58) Shakin-Eshleman S. H., S. L. Spitalnik & L. Kasturi, (1996) The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. *J Biol Chem* 271: 6363-6366.
- (59) Holst, B., Bruun, A. W., Kielland-Brandt, M. C. and Winther, J. R. (1996). Competition between folding and glycosylation in the endoplasmic reticulum. *EMBO J*. 15, 3538-3546.
- (60) Malkin LI, Rich A. Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *J Mol Biol*. 1967 Jun 14;26(2):329-46.
- (61) D Sabatini; G Blobel. Controlled proteolysis of nascent polypeptides in rat liver cell fractions. II. Location of the polypeptides in rough microsomes. *The Journal of cell biology* 1970;45(1):146-57.
- (62) G Blobel; D D Sabatini. Controlled proteolysis of nascent polypeptides in rat liver cell fractions. I. Location of the polypeptides within ribosomes. *The Journal of cell biology* 1970;45(1):130-45.
- (63) Whitley P, Nilsson IM, von Heijne G. A nascent secretory protein may traverse the ribosome/endoplasmic reticulum translocase complex as an extended chain. *J Biol Chem*. 1996 Mar 15;271(11):6241-4.
- (64) Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*. 2007 Jan 26;315(5811):525-8. Epub 2006 Dec 21.
- (65) Zhang G, Ignatova Z. Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Curr Opin Struct Biol*. 2011 Feb;21(1):25-31. doi: 10.1016/j.sbi.2010.10.008. Epub 2010 Nov 24.
- (66) Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol*. 2009 Mar;16(3):274-80. doi: 10.1038/nsmb.1554. Epub 2009 Feb 8.
- (67) Clarke TF 4th, Clark PL. Rare codons cluster. *PLoS One*. 2008;3(10):e3412. doi: 10.1371/journal.pone.0003412. Epub 2008 Oct 15.
- (68) Zhang F, Saha S, Shabalina SA, Kashina A. (2010) Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation. *Science* 329:1534-7.
- (69) Walmsley AR and Hooper NM. Distance of sequons to the C-terminus influences the cellular N-glycosylation of the protein. *Biochem. J.* (2003) 370, 351-355

(70) I. Nilsson, G. von Heijne, Glycosylation efficiency of Asn-Xaa-Thr sequons depends both on the distance from the C terminus and on the presence of a downstream transmembrane segment, *J. Biol. Chem.* 275 (2000) 17338–17343

(71) Jones J, Krag SS, Betenbaugh MJ. Controlling N-linked glycan site occupancy. *Biochim Biophys Acta*. 2005 Nov 15;1726(2):121-37. Epub 2005 Jul 25.

(72) Rodriguez J, Spearman M, Tharmalingam T, Sunley K, Lodewyckx C, Huzel N, Butler M. High productivity of human recombinant beta-interferon from a low-temperature perfusion culture. *J Biotechnol*. 2010 Dec;150(4):509-18. doi: 10.1016/j.jbiotec.2010.09.959. Epub 2010 Oct 8.

(73) Nilsson I, von Heijne G. Glycosylation efficiency of Asn-Xaa-Thr sequons depends both on the distance from the C terminus and on the presence of a downstream transmembrane segment. *J Biol Chem*. 2000;275:17338–1734

(74) Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* 24: 79-88, 2006.

(75) B.L. Schulz, C.U. Stirnimann, J.P. Grimshaw, M.S. Brozzo, F. Fritsch, E. Mohorko, G. Capitani, R. Glockshuber, M.G. Grutter, M. Aebi. Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl. Acad. Sci. U. S. A.*, 106 (2009), pp. 11061-11066.

(76) Suzuki, T., Arai, S., Takeuchi, M., Sakurai, C., Eabana, H., Higashi, T., Hashimoto, H., Hatsuzawa, K., and Wada, I.: Development of cysteine-free fluorescent protein for the oxidative environment, *PLoS One*, 7, e37551 (2012).

(77) López Medus M, Gomez GE, Zacchi LF, et al. N-glycosylation Triggers a Dual Selection Pressure in Eukaryotic Secretory Proteins. *Scientific Reports*. 2017;7:8788.

(78) J. L. Mellquist, L. Kasturi, S. L. Spitalnik and S. H. Shakin-Eshleman. The Amino Acid Following an Asn-X-Ser/Thr Sequon Is an Important Determinant of N-Linked Core Glycosylation Efficiency. *Biochemistry* **1998**, 37, 6833-6837

(79) Cherepanova NA, Gilmore R. Mammalian cells lacking either the cotranslational or posttranslational oligosaccharyltransferase complex display substrate-dependent defects in asparagine linked glycosylation. *Scientific reports*. 2016;6:20946

(80) Malaby, H. L. and Kobertz, W. R. (2013) Molecular determinants of co- and post-translational N-glycosylation of type I transmembrane peptides *Biochem. J.* 453, 427– 434

(81) Iben JR, Maraia RJ (2014) tRNA gene copy number variation in humans. *Gene* 536:376–384.

(82) PETER J. NIELSEN AND EDWIN H. McCONKEY Evidence for Control of Protein Synthesis in HeLa Cells via the Elongation Rate JOURNAL OF CELLULAR PHYSIOLOGY 104:269-281 (1980).

(83) Pfeffer, Stefan, et al. "Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon." Nature communications 5 (2014): 3072.

(84) Roboti, P., and S. High. 2012b. The oligosaccharyltransferase subunits OST48, DAD1 and KCP2 function as ubiquitous and selective modulators of mammalian N-glycosylation. J. Cell Sci. 125:3474–3484.

(85) S. Shrimal, N. A. Cherepanova, R. Gilmore, DC2 and KCP2 mediate the interaction between the oligosaccharyltransferase and the ER translocon. J. Cell Biol. 213, 3625-3638 (2017)

(86) Shrimal, S., Trueman, S. F. and Gilmore, R. (2013b). Extreme C-terminal sites are posttranslocationally glycosylated by the STT3B isoform of the OST. J. Cell Biol. 201, 81–95.

(87) Solá, R.J., Rodriguez-Martinez, J.A., Griebenow, K. Modulation of protein biophysical properties by chemical glycosylation: Biochemical insights and biomedical implications. Cell Mol Life Sci. 2007; 64:2133–2152.

(88) Wang, C., et al. (1996) Influence of the carbohydrate moiety on the stability of glycoproteins. Biochemistry, 35, 7299–7307.